# Time Required to Classify Different Sound Types and Their Relation to Hearing Ability and Speech Understanding in Noise

A thesis submitted in partial fulfilment of the requirements of the degree of

Master of Audiology

By Durga Lal Budathoki

Department of Communication Disorders

University of Canterbury

2019

## Abstract

It is well known that people with sensorineural hearing loss (SNHL) have difficulties with basic auditory processing abilities such as temporal processing, frequency selectivity, frequency discrimination, as well as difficulties with complex abilities like speech perception in noise, music perception, and environmental sound awareness. The study aim of the study was to explore the procedures to reliably measure the minimum duration of sounds needed to classify the four major categories of sounds. The minimum duration of sounds needed to classify sounds was termed classification time thresholds. Another aim of the study was to investigate whether the classification time thresholds would correlate with pure tone average (PTA) thresholds. In addition, this study aimed to investigate whether the classification time thresholds would correlate with understanding of speech in noise. The classification time thresholds for classifying speech, music, noise, and animal sounds were measured in participants with normal-hearing and those with SNHL, using an adaptive 4-AFC procedure. In addition, the participants underwent pure-tone audiometry and speech in noise testing using the New Zealand matrix sentence test in auditory-alone mode.

The study showed that the participants with SNHL took longer to classify speech, noise, and music sounds. This may be due to impairment in processing abilities like temporal resolution, perception of temporal fine structure, frequency selectivity, and frequency discriminations. The study also showed that the better the ability in classifying short speech, noise, and music sounds, the better the understanding of speech in noise. This finding is consistent with glimpsing model of speech understanding in noise. Hearing ability was correlated to ageing, and to classification time thresholds. Both the effect of aging and hearing loss may cause deficits in abilities required for classifying short sounds. However, this study was unable to separate the independent effect of age and SNHL on classification time thresholds. This study may serve as an initial step towards reliably measuring classification time thresholds for participants with hearing loss and normal hearing. This study has documented the procedures that are effective and those that are problematic.

## Acknowledgements

I would like to thank Associate Professor Greg O'Beirne of the University of Canterbury for supporting me in bringing this work to fruition. I would also like to thank Professor Jürgen Tchorz of the Lübeck University of Applied Sciences for allowing me to expand on his work and helping me with the statistical analysis.

I wish to express my gratitude to staff of Audiology department at the University of Canterbury, and the amazing classmates for providing a supportive environment.

Thank you to my parents and my four amazing sisters: Durga, Devi, Chandra, and Lalita for their unwavering support and a reason to excel. Thank you to my steadfast star Sumanti for all the love and patience, throughout my study in Christchurch.

## Table of Contents

## List of Figures

## List of Tables

## List of Abbreviation

| | |
|---|---|
| **4-AFC** | Four-alternative forced choice method |
| **AC** | Air conduction |
| **BC** | Bone conduction |
| **BM** | Basilar membrane |
| **dB A** | A-weighted decibels |
| **dB HL** | Decibel Hearing level |
| **dB SPL** | Decibel sound pressure level |
| **EEG** | Electroencephalography |
| **FDL** | Frequency difference limen |
| **FMDL** | Frequency modulation detection limen |
| **HFPTA** | High frequency PTA |
| **HI** | Hearing impairment |
| **HINT** | Hearing in Noise Test |
| **Hz** | Hertz |
| **IHCs** | Inner hair cells |
| **kHz** | Kilo Hertz |
| **ms** | millisecond |
| **MST** | Matrix sentence test |
| **NH** | Normal hearing |
| **OHCs** | Outer hair cells |
| **PTA** | Pure tone average |
| **PTC** | Psychophysical tuning curves |
| **QuickSIN** | Quick Speech-in-Noise |
| **SL** | Sensational level |
| **SNHL** | Sensorineural hearing loss |

| | |
|---|---|
| **SNR** | Signal-to-noise ratio |
| **SPIN** | Speech in Noise Test |
| **SRT** | Speech reception threshold |
| **SRT** | Speech reception in noise thresholds |
| **TFS** | Temporal fine structure |
| **TI** | Temporal integrator |
| **TMTF** | Temporal modulated transfer function |
| **TX** | Texas |
| **U.S.A** | United States of America |
| **UCAMST** | University of Canterbury Auditory-visual Matrix Sentence Test |
| **UCSHC** | University of Canterbury, Speech and Hearing Clinic |

# 1. Introduction

## 1.1 Overview of hearing loss and its impact

Hearing loss is the most common sensory deficit. The World Health Organisation reported that it impacts 360 million people worldwide (Duthey, 2013). Data collected from 17 district health boards across New Zealand showed that the most common type of hearing loss is sensorineural hearing loss (SNHL) at around 60 percent, followed by mixed hearing loss at 10 percent, and permanent conductive hearing loss at around percent (Digby, Purdy, & Kelly, 2016). Hearing loss impacts a person's ability to carry out conversations. It also reduces the quality of life and increases self-perceived handicap (Dalton, Cruickshanks, Barbara, & Klein, 2003). Furthermore, it prevents people from participating in social activities. Hearing loss can seriously affect the mental health of a person (Tambs, 2004). It can also impact the people surrounding a person with the impairment (Scarinci, Worrall, & Hickson, 2008). Elderly are most likely to have hearing loss compared to all other age groups. Half of everyone above 75 years old has hearing loss and so does a third of the population between 65 and 75 years old (National Institute of Heath, 2019). The hearing loss associated with old age is called presbycusis is one form of SNHL .

## 1.2 Overview of anatomy and physiology of hearing

The peripheral auditory anatomy is divided into three sections: outer ear, middle ear, and the inner ear. The outer ear consists of the pinna (visible ear), the auditory meatus (ear canal) and the eardrum (tympanic membrane). One of the functions of the pinna is to maximise the sound pressure level (SPL) arriving at the ear drum. Another function is to aid the localisation of sound (Rice, May, Spirou, & Young, 1992). The middle ear consists of three tiny bones called the ossicles (individually called malleus, incus, and stapes). The chain of middle ear bones joins the outer ear and the inner ear, and its function is impendence matching. The sound travels through medium of air in the outer ear but through fluid in the inner ear. If the airborne sound waves were to directly hit the surface of the inner ear, a significant amount of sound would be reflected instead of propagating into the inner ear. The pressure at eardrum is increased by almost 200 times by the time sound waves reach the inner ear due to middle ear function in impedance matching. The impedance matching is achieved through the two ways. Firstly, by focusing the movement at the large diameter of tympanic membrane to a small diameter of oval window (where the ear bones contact the inner ear). Secondly, by the mechanical lever action of the three middle ear bones (Aibara, Welsh, Puria, & Goode, 2001). Any type of dysfunction in the outer ear or the middle ear that results in the impairment of hearing is called conductive hearing loss (Zwicker & Schorn, 1978).

The inner ear consists of vestibular labyrinth and the cochlea. The main function of vestibular labyrinth is to provide sensory information regarding the perception of rotational and linear motion of the head in relation to the gravity (Ekdale, 2016). The cochlea is the hearing organ where the mechanical signals of the sound waves are turned into electrical impulses. This process is called neural transduction. The cochlea is a snail shaped fluid filled organ which houses cells required for neural transduction (Ruggero & Temchin, 2002). The sensory hair cells in the cochlea are located on the basilar membrane (BM). The BM is tonotopically organised, which means that the low frequency sound waves cause vibrations at the apical end and the high frequency cause vibrations at the basal end. The location of maximum displacement of BM, is frequency specific (Stasiunas et al., 2003).

There are two types of hair cells in the BM, which are called inner hair cells (IHCs) and outer hair cells (OHCs). There are approximately three times as many OHCs as IHCs. The function of the inner hair cell is to turn the mechanical energy of the BM displacement and bending of stereocilia into electrical signal. When the BM vibrates, it bends the stereocilia that sits on top of the BM. The bending of stereocilia opens a mechanically gated ion channel, which triggers the process for generation of electrical signal called neural transduction. This electrical signal travels to the brain via the afferent nerve connecting the IHCs (Pickles, 2012). The process of sound energy causing BM vibration which triggers neural transduction without the involvement of OHCs is called a passive process (Stasiunas et al., 2003). The OHCs are connected to the membrane called tectorial membrane directly above them. The function of the OHCs is to amplify the vibration of the BM. The role of OHCs is also associated with refining sensitivity and frequency selectivity of the cochlea (Stasiunas et al, 2003; Ashmore & Kolston, 1994). The mechanical amplification is thought to be carried out by a motor protein called "prestin" in the basal membrane of the OHCs. The higher the intensity of BM vibration, the higher the number of hair cells firing, and the louder the sound perceived is. This is called active mechanism. The amplification of BM vibration caused by OHCs is thought to directly contribute to the non-linear growth of the cochlea response (Pickles, 2012).

Any type of dysfunction in the cochlea or in the peripheral auditory system after the cochlea that results in hearing impairment is called sensorineural hearing loss (SNHL). There are two types of sensorineural hearing loss: retrocochlear and cochlear. They are defined by the location of the underlying impairment. Hearing loss as a result of damage beyond the cochlea is called retrocochlear SNHL (Patuzzi, 2009). Whereas, cochlear hearing loss occurs as a result of the damage or dysfunction of the OHCs or the IHCs in the cochlea. Damage to OHCs are called motor hearing loss and damage to IHCs are called sensory hearing loss. If there is damage to OHCs (motor hearing), the active process is impaired. This makes the cochlea more dependent on the passive process. Damage to the active process is also thought to impair the function of the peripheral auditory system called compressive non-linearity (Moore, 2013). See section 1.3 for details on compressive non-linearity. The impact of SNHL is more damaging compared to conductive hearing loss because it impairs both the sensitivity and the frequency specificity. This means that not only the intensity of sound required for threshold of hearing increases (sensitivity) but the ability to clearly understand speech also diminishes. The clarity of speech provided by high functioning frequency selectivity diminishes (Patuzzi, 2009).

There are different causes of sensorineural hearing loss/impairment such as aging (presbycusis), ototoxic drugs, head trauma, noise exposure, and hereditary conditions. The cochlear OHCs and IHCs, and afferent neurons have very limited capacity to repair or regenerate themselves. This makes sensorineural hearing loss permanent in most cases, including cases where the aetiology is presbycusis or hereditary (Wong & Ryan, 2015). Sensorineural hearing impairment can impact important cochlear functions like compressive nonlinearity (Moore, 2009). As well as basic signal processing tasks such as temporal resolution, frequency selectivity and frequency discrimination, and high-level processing like understanding speech in noise.

### 1.3 Compressive nonlinearity and hearing loss

As mentioned above, one of the functions affected by sensorineural hearing loss resulting from damage to OHCs is compressive nonlinearity. Compressive nonlinearity allows a wide range of sound level input to be represented by a small range of BM movement. The intensity of the compression is not constant across all levels of sound input. For low input sounds, below 20 dB SPL, the function is linear. As well as for the very high intensity sound, above 90 dB SPL. However, for mid-range levels, the slope is less steep. The active process plays a vital role in the nonlinear response of cochlea. This is because the active mechanism increases the response of the BM for low and mid-level sounds. According to Moore (2013), the gain or amplification provided by the active mechanism may be 50 dB or more. At low input level, below 20 to 30 dB SPL, the gain is at its peak and constant. The amplification progressively decreases as the input level increases. Therefore, at midlevel, the growth is nonlinear. However, when the input level reaches 90 dB SPL, there is no contribution from active process. This makes the response linear once again.

The compression occurs maximally at the peak of the BM vibration. For example, if the input frequency was a 2 kHz sinusoid wave, the maximum displacement in the BM and maximum compression would only occur at the place on the BM corresponding to 2 KHz. The nonlinear function decreases with the degradation of physiology of the cochlea (Cooper & Rhode, 1995). Sensorineural hearing loss results in less compressive input-output function. The less compressive input-output function causes the cochlea to act in a more linear way. This results in hearing impaired people perceiving abnormally rapid growth in loudness with increasing sound pressure level. This can lead to extreme discomfort in some cases. The process of abnormal growth of perceived loudness is called loudness recruitment. The damage to compressive nonlinearity is also thought to impact temporal processing (Moore, 2013).

**1.4 Temporal Resolution**

Temporal resolution is defined as the ability of an individual to detect differences in auditory stimuli duration and in interval between stimuli over time. Temporal resolution describes the resolution of changes in envelope rather than the fine temporal structure. The temporal fine structure (TFS) is the rapid changes in the amplitudes due to the rapid changes in the sound pressure. Whereas, envelope is the slower changes in the amplitude fluctuations but superimposed within temporal fine structure as shown in Figure 1.1.

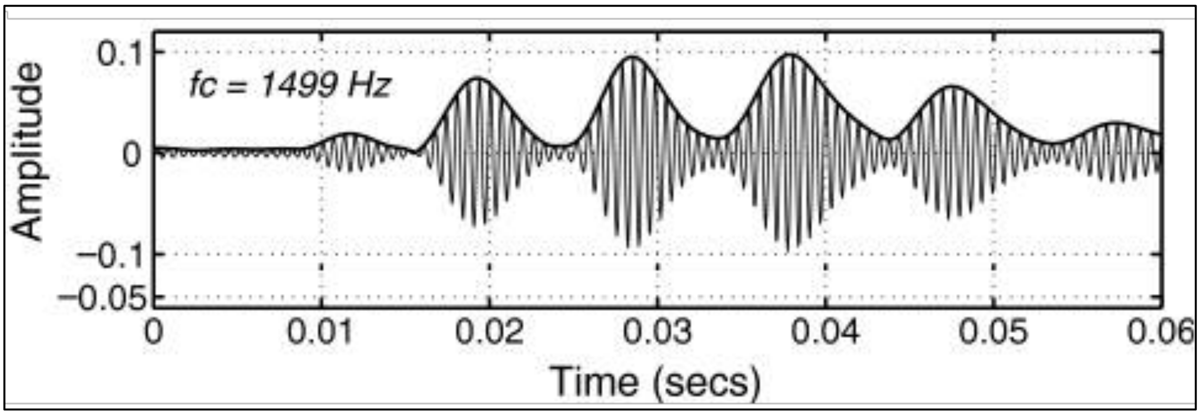


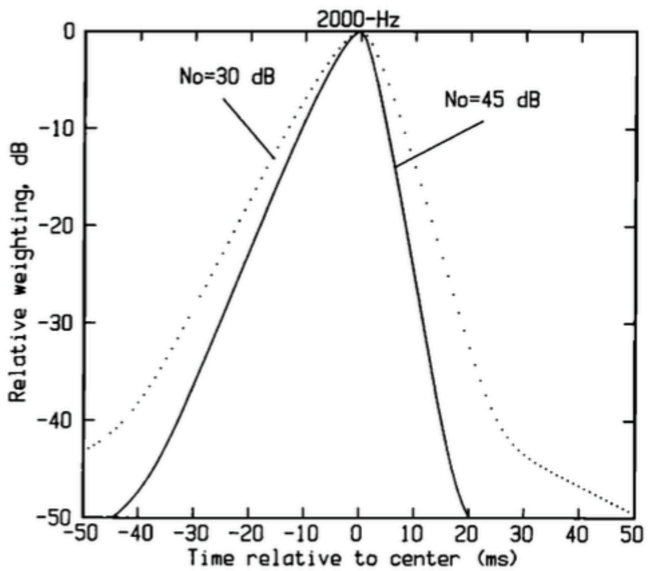
*Figure 1.1.* The difference between envelope and temporal fine structure. The darker line on the top shows the slow fluctuations in the amplitude known as envelope. The thinner line with rapid fluctuations in amplitude shows temporal fine structure. Adapted from Moore (2008).

### 1.4.1 A Model of temporal processing

One way of modelling temporal processing is described as a hierarchical system consisting of numerous processing steps. In the description below, the use of the term “device” is referred to any site where this processing takes place. In short, the stimulus arrives at the basilar membrane (BM) and then passes through a band-pass filter, then through to nonlinear device, to sliding temporal integrator, and then to decision device.

The peripheral auditory system, specifically the basilar membrane, is thought to contain a series of band-pass filters with overlapping pass-bands called auditory filters. It is thought that each section of basilar membrane is responsible for signal transduction of certain frequencies. So each sections of basilar membrane has filters with different centre frequencies, that overlaps with the adjacent filter (Fletcher, 1940; Moore & Glasberg, 1986). The filter’s shape can be estimated using different techniques, which are described in greater detail in Section 1.6.1, when discussing frequency selectivity. The auditory filter influences temporal processing at mainly low centre frequencies below 200 Hz but at higher centre frequency, higher processing centre influence the temporal processing (Moore, 2008). The non-linear device succeeds each filter. Filtering and non-linear process cannot be viewed as separate stages in realistic manner, but it does not affect the model to do so for simplicity. It reflects several processes of peripheral auditory system including half-wave rectification and compressive input-output function of basilar membrane (Moore, 2008). Compressive input-output function or compressive non-linearity is described in Section 1.3.

The sliding temporal integrator (TI) is a device in which the output of non-linear device becomes the input and where it performs ‘smoothing’. The output of the TI is the weighted running average of the output of non-linear device that happens across specific time interval or window. The time window is called the shape of the temporal window, as shown in Figure 1.2 below. This results in the output of TI smoothing the rapid fluctuations, whereas the slow fluctuations are not affected (Moore, 2008). The TI takes time to build up and fade-in when there is an abrupt presentation or turn off signal. The presence of temporal integrator can be explained by the phenomena of forward and backward masking. In these cases, non-simultaneous presentation of a signal, before or after a masking noise, can have an impact on the detectability of the signal due to build up and fade in of TI introducing a masking effect (Oxenham & Moore, 1994; Moore, 2013).




.

*Figure 1.2.* An example of the shape of the temporal window derived from 2000Hz signal of 5ms duration showing that more weighting is given to the temporal centre. Also shows that the slope to the right of centre is steeper compared to the left. Adapted from Moore, Glasberg, Plack, & Biswas (1988).

The sliding TI is followed by a decision device that follows a specific set of rules already established to carry out specific tasks. This includes separate rules for gap-detection and amplitude modulation (Moore, 2008). Temporal characteristics of speech, like brief silent intervals and rapid modulation of intensity, provide perceptual cues for recognition of phoneme, word sentences, as well as prosodic information (John, Hall, & Kreisman, 2012).

### 1.4.2 Relationship between temporal resolution, hearing ability and speech understanding in noise.

Two main tests that measure temporal resolution are gap detection tests and temporal modulated transfer function (TMTF) tests (Moore, 2013). Gap detection threshold is measured as the shortest silent period within a tone or burst of noise that an individual can detect. Whereas, the TMTF measures the function between the modulation rate of a sinusoid wave and an individual's ability to detect the amplitude modulations (Moore, 2013). Multiple studies have shown that performance in temporal resolution tests and perception of degraded or distorted speech are associated with each other (Gordon-Salant & Fitzgibbons, 1999; Schneider & Pichora-Fuller, 2001; Snell, Mapes, Hickman, & Frisina, 2002; Pichora-Fuller, Schneider, MacDonald, Pass, & Brown, 2007). Studies have also shown that the gap detection in noise and tones is correlated to speech perception in noise and reverberation (Snell et al., 2002; Tyler, Summerfield, Wood, & Fernandes, 1982; Gordon-Salant & Fitzgibbons, 1993).

The performance in gap detection tests varies amongst people with similar audiograms and is poorly predicted by the audiogram (Glasberg, Moore, & Bacon, 1987; Schneider, Pichora-Fuller, Kowalchuk, & Lamb, 1994). When controlling for audibility, some have found association between speech perception and gap detection thresholds (Phillips, Gordon-Salant Fitzgibbons, & Yeni-Komshian, 2000; Haubert & Pichora-Fuller, 1999), others have not (Snell & Frisina, 2000; Strouse, Ashmead, Ohde, & Grantham, 1998). This differing results might be due to difference in stimuli and procedures (John, Hall, & Kreisman, 2012). Factors such as age are also thought to affect temporal processing, with older adults performing worse at gap detection tests (Lister, Roberts, & Lister, 2011; Snell, 1997).

Glasberg, Moore, & Bacon, (1987), used the rate of decay of forward masking to measure temporal resolution. They found that at same sound pressure level the individuals with sensorineural hearing impairment had more rapid decay of forward masking, indicating poor temporal resolution. Whereas, at same sensational level (SL) the difference between the two group was reduced. The performance of HI individuals was worse at low SL, and increasing it resulted in discomfort due to loudness recruitment. Damage to the outer hair cells of the cochlea results in loudness recruitment. Less compressive input-output function of the basilar membrane causes the cochlea to act in a linear way, resulting in hearing impaired people perceiving abnormally rapid growth in loudness with increasing sound pressure level. Due to the hearing impairment, these people listen at low SL in daily life. The study concluded that the temporal resolution was poor for hearing impaired in comparison to those with normal hearing (Moore, 2008).

For some aspects of temporal resolution, even when the audibility is controlled, people with HI struggle. For example, Florentine and Buss (1984) showed that the gap detection threshold is higher for people with HI when there is slow and random fluctuation in amplitude like in noise with narrow bandwidth. This might be due to loudness recruitment because recruitment causes the inherent amplitude fluctuations of noise to be heard as abnormally large fluctuations in loudness. This can be confused with the gaps in the noise, hindering their ability for gap detection (Moore, Wojtczak & Vickers, 1996). This finding is supported by data from Glasberg and Moore (1992), where they showed that stimulating recruitment in the normal ear produced increased gap detection thresholds like those found in HI participants. Most sounds in everyday life contain slow random fluctuations from one movement to other. This along with decreased sensational levels hinders the ability of people with HI to follow the temporal architecture of the sound leading to decreased understanding of speech, identifying music and environmental sounds (Moore, 2008).

### 1.5 Frequency discrimination and hearing ability and speech perception

HI individuals also have difficulty with frequency discrimination. Frequency discrimination is known as the ability to distinguish the changes in frequency over time (Moore, 2008). The physical changes in frequency is heard as changes in perceptual pitch. Difference limen is the smallest detectable change in frequency. Frequency discrimination can be measured using two methods; frequency difference limen (FDL) and frequency modulation detection limen (FMDL). Multiple studies have found that FDL is affected by sensorineural hearing impairment (Moore, Glasberg, & Peters, 1996; Hall & Wood, 1984; Simon & Yund, 1993). There is also a great degree of variability between individuals as FDL is not predicted by pure-tone thresholds. It has also been found that FDL in two ears of the same individual are different even when the pure tone thresholds are identical (Simon & Yund, 1993). FMDLs have also been shown to be larger in hearing impaired individuals indicating worse frequency discrimination (Grant, 1987; Zurek & Formby, 1981). Frequency discrimination plays a major role in speech perception. It is necessary in determining prosody of a language. In tonal languages, the change in pitch can change the meaning of words. The ability to hear the voice of just one person in a crowded place also relies on the having good pitch perception. It allows listener to focus on the fundamental frequency of the speaker, as voices of different people have different fundamental frequency (Brokx & Nooteboom, 1982).

### 1.6 Frequency selectivity and auditory filters

Frequency selectivity which is different to frequency discrimination is also affected in person with sensorineural hearing impairment. Frequency selectivity, which is also known as frequency resolution, is an ability to separate or resolve different sinusoidal components in a complex sound (Moore, 2013). Frequency selectivity can be measured using masking. The two of the methods are psychophysical tuning curves (PTC) and notched noise methods. Both methods estimate the shape of the auditory filters based on assumptions of the power spectrum model. The power spectrum model assumes that the listener uses an auditory filter with centre frequency close to the signal, when detecting signal in noisy environment. This means that the filter allows the signal to pass but removes a lot of noise (Fletcher, 1940). Only the parts of the noise that pass through the filter being used will have any effect on masking. It is also assumed that the amount of noise that pass through the filter determines the threshold for a signal. There is a certain signal-to-noise ratio at the output of the filter that corresponds to the threshold for a signal (Patterson & Moore, 1986).

### 1.6.1 Relationship between hearing ability, frequency selectivity and auditory filters

Estimation of auditory filters in normal and hearing impaired using various techniques by several studies have showed that the auditory filters are significantly broader in people with sensorineural hearing impairment in comparison to people with normal hearing (Florentine, Buss, Scharf, & Zwicker, 1980; Glasberg & Moore, 1986; Pick, Evan, & Wilson, 1977). Measures of frequency selectivity on people with different types of hearing loss show that people with sensorineural loss have broader auditory filters in comparison to those with conductive hearing impairment indicating that the loss of frequency selectivity was associated with sensorineural loss (Zwicker & Schorn 1978). There is a weak correlation between pure tone thresholds and broadening of the filters, but higher thresholds tend to be associated with broader filters (Tyler, 1986). Moore (2013), reported that the broadening of the filters mainly depends on the OHC damage.

Impaired frequency selectivity interferes with perceptual abilities such as greater masking of signal by background sounds as broader filters means that more noises pass through it. Listeners utilise auditory filters with best signal to noise ratio when detecting signal in noisy background. In normal hearing person, the auditory filter is narrow. Most of the noise does not pass through the filter with the signal, producing better signal-to-noise ratio at the filter output. Whereas, in person with HI, the filter being used is broader. This allows lots of noise to pass along with the signal resulting in reduced signal detectability (Moore, 2013). Therefore, in the presence of background noise, the ability to detect and discriminate the signals are poorer in person with sensorineural hearing impairment this includes speech and music. The reduced ability for frequency selection may partially account for reduced ability of people with sensorineural hearing impairment to understand speech in background noise (Moore, 2008).

Frequency selection is also required for perception of timbre of music. The ability to detect differences in spectral composition of sound is required for discrimination between different vowel sounds and different musical instruments. When frequency selection is reduced, this ability is also reduced. This leads to difficulty in discriminating between different vowel sounds as well as different musical instruments. Providing amplification with hearing aids increases the audibility of signals, but they cannot fix impaired frequency selectivity (Moore 2013).

**1.7 Hearing impairment and speech understanding in noise.**

The primary concern of hearing-impaired people visiting the clinic is difficulty with understanding speech in noise (Kochkin, 2002). People with hearing loss find it hard to understand speech in presence of background noise. A person's ability to understand speech in noise can be measured using speech in noise tests. It is defined by the value called speech reception threshold (SRT) in noise (McArdle & Hnath-Chisolm, 2015). There are multiple tests that measure SRT in noise. Some of the widely used tests are, Speech in Noise (SPIN), Hearing in Noise Test (HINT), Quick Speech-in-Noise (QuickSIN) (Kalikow, Stevens, &

Elliott, 1977; Nilsson, Soli, & Sullivan, 1994; Killion, Niquette, Gudmundsen, Revit, & Banerjee, 2004). The main goal of these tests is to measure the signal-to-noise ratio (SNR) required to understand certain percentage of speech signal in background noise. For HINT and QuickSIN along with other tests, SRT is defined as the SNR required to understand 50% of the speech sounds. The type of noise and speech stimuli used are different for each test and they each have their own advantages and disadvantages (McArdle & Hnath-Chisolm, 2015).

The noise used can either be held at a constant level or can be fluctuating like multiple-talker babble noise. Tests with constant level of noise usually use narrow band noise, that can effectively mask speech sounds. They show that it produces results that have higher sensitivity in comparison to tests using fluctuating noise (Wagener & Brand, 2005). The multi-talker babble noise is thought to better represent everyday noises like noises in cafes, restaurants, and crowds. However, the inherent fluctuations of the SNR in the multi-taker babble is thought to affect accuracy of SRT being measured (Killion, et al., 2004; Hochmuth, Kollmeier, Brand, & Jürgens, 2015).

The speech stimuli can either be a monosyllabic word or a sentence. The use of monosyllabic word provides little contextual cues. Unlike sentence stimuli, the auditory memory does not influence the results. Tests using sentences more closely resemble everyday experiences. Therefore, they provide higher face validity compared to tests with monosyllabic word (Wilson, McArdle, & Smith, 2007). The tests with sentences stimuli measure multiple speech sounds per trial, which also increases time-efficiency. However, unimpaired auditory memory is required to remember the complete sentences to be able to repeat them during each trials (Hochmuth et al., 2012).

There are two main types of sentence test: “Plomp type” sentence test or Matrix sentence test (Wagener, Brand, & Kollmeier, 2007). “Plomp type” sentence tests are phonetically balanced, short, and meaningful sentences, based on everyday conversation. However, the grammatical structure of the sentences is not consistent across the trials. HINT is one of the widely used tests that use “Plomp type” sentences consisting of 25 phonetically sentences (Plomp & Mimpen, 1979).

Matrix sentence test (MST) was first developed by Hagerman (1982). MST sentences consist of 10 five-word sentences with the following structure: name, verb, number, adjective, and object. For example, “David has ten old toys”. The grammatical structure is consistent and allows for a high number of unique sentences to be created from a group of few words, when they are switched and replaced. MST has a few other advantages over other sentence test. It has low redundancy because the number of possible sentences that can be used are very high. In addition, it is semantically unpredictable and the listener receives low contextual cues (Hochmuth, Kollmeier, Brand, & Jürgen, 2015). MST sentences can be easily changed to accommodate for different languages. MST is available in many different languages including, Italian (Puglisi et al. 2014), Spanish (Hochmuth et al., 2012), Polish (Ozimek, Warzybok, & Kutzner, 2010), and New Zealand English (Trounson, 2012).

Only 3% of clinicians, surveyed by Mueller (2003), used formal speech in noise testing. The most common test of hearing is done by using pure-tone thresholds in quiet enviroment. Pure tones are the simplest form of sound. The two important characteristics of pure tones in clinical audiology are frequency and amplitude (sound level) measured in Hertz (Hz) and decibel (dB) HL respectively. Pure tone audiometry usually measure frequency between 250 Hz and 8000 Hz which is close to the range (100 - 6000 Hz) required for understanding speech (French & Steinberg, 1974). Pure tone thresholds are the lowest level of sound required to elicit a response. The pure tone thresholds show quantifiable frequency-specific hearing impairments for each ear. They are very important in diagnosing hearing loss and its source (Schlauch & Nelson, 2015).

However, pure tone thresholds do not predict the performance of an individual in real world like speech perception in noise does. Some studies claimed that there is no significant relationship between pure tone thresholds and speech-in-noise understanding (Duquesnoy, 1983; Blandy & Lutman 2005; Middelweerd, Festen, & Plomp, 1990). However, other studies like Smoorenburg (1992), and Lutman (1991) claimed that there is a relationship between pure tone thresholds and speech understating in noise, and reported that as the pure tone average thresholds increases the understanding of speech in noise decreases. The performance of people with similar pure-tone thresholds in speech understanding in noise greatly varies. Similar speech in noise performance may also be found in people with very different pure-tone thresholds (Blandy & Lutman, 2005; Smoorenburg, 1992). Turner, Fabry, Barrett, and Horwitz, (1992) reported that people with hearing impairment required greater signal-to-noise ratio to recognise stop consonants embedded in white noise in comparison to individuals with normal hearing even though the detection occurred at the same level for both groups. The loss of audibility can be restored by hearing aids. However, the ability of hearing aids to help understand speech in noise is limited. The newer hearing aids provide better signal to noise ratio and are better than older hearing aids at aiding speech perception in noise but they are incomparable to the abilities of normal ears (Healy & Yoho, 2016).

One of the theoretical mechanisms for speech recognition in noise is called the glimpsing model. According to this model, the listeners extract information from parts time-frequency which have clear signal and relatively better signal to noise ratio. Speech perception in noise is based on the ability of the listener to perform extractions from these brief clear parts (Kidd & Humes, 2012). One of the major determinants of speech understanding in noise is thought to be the proportion of glimpses performed on the signal by the listener (Kidd & Humes, 2012). Thus, the ability to distinguish speech in noise requires the ability to accurately and rapidly separate speech signals from noise in those brief moments of better signal to noise ratio (Cooke, 2006). Therefore, it could be said that having a better temporal resolution in identifying or categorising brief sounds is important for understanding speech in noise. The ability to identify sounds in shorter time should help better in understanding speech in noise.

**1.8 Factors involved in identification of short sounds.**

Many factors are involved in the identification of short sounds including, acoustics, perceptual and cognitive factors (Ballas, 1993). They can either be external or internal factors. Acoustics characteristics of the sound clips are an external factor. Perceptual and cognitive factors are internal factors.

One of the external acoustic factors that influence identification is spectral resolution of the sound. Shafiro (2008) used 60 environmental sounds, such as mechanical sounds, water sounds, sounds made by humans and animals among others. Limiting the frequency range of the sounds between 500 Hz and 5 kHz with the preserved fine spectral structure had no significant impact in the accuracy of identification of the sounds. However, it reported that decreasing the spectral resolution decreased the accuracy of identification of environmental sounds. Turner, Chi, and Flock (1999) also showed that limiting spectral resolution produced poorer performance in recognising speech consonants in participants with moderate sensorineural hearing-impaired loss in comparison to participants with normal hearing.

Ogg, Slevc, and Idsardi, (2017) investigated the gated thresholds required to classify speech, music and environmental sounds in participants who reported to have normal hearing. They showed that the variation in the fundamental frequency was an important cue in identifying music (instrumental) and speech of short durations. The participants used the variation in the pitch to differentiate between speech and environmental sounds. The nosiness of sound was associated with environmental sounds and spectral flatness was associated with instruments. Model based on responses of participants indicated that the spectral centroid was related to response for all classes of sound. Ogg et al., (2017, p.3463) defines spectral centroid as “Spectral centre of gravity corresponding to the mean of the amplitude weighted spectrum”. When the duration of sounds were below 75ms, speech was more confused with environmental sounds when consonants were at the beginning of sounds compared to when a vowels were at the start.

There are also number of perceptual and cognitive factors associated with identification of short sounds. Ballas (1993) tested the time required to correctly identify brief everyday sounds in normal hearing participants. In a post testing questionnaire, the participants rated each sound for familiarity. The study found a significant correlation between self-reported familiarity and the time taken to identify the sound. Familiar sounds took shorter time to identify compared to less familiar sounds. Shafiro (2008) has also shown that accuracy of identification is correlated to the familiarity of sounds. These findings are consistent with the finding of Cycowicz and Friedman (1998), which used event-related evoked potentials to show that familiarity plays an important role in the identification of environmental sounds.

Ballas (1993) also reported that there is a relationship between the identifiably of sound and the ease of forming a mental picture of the source of the sound. In another experiment, participants were asked to listen to different sounds with a probable cause simultaneously displayed on the screen. Participants were told to indicate whether the sound they heard matched the description of the cause. The identification time for the sounds was faster if the cause was more probable compared to when the cause was less probable cause (Ballas, 1993).

Giordano, McDonnell, and McAdams (2010) investigated how sounds created by a living or a non-living source were processed. They reported that sounds from non-living sources were identified using a robust strategy that was independent of context and heavily focused on acoustical information. Whereas, sounds from living sources were identified using flexible cognitive strategy that used symbolic information as well as acoustical information and was heavily dependent on the context.

Attention and memory are important cognitive factors involved in auditory perception (Chiu & Schacter, 1995; Shen, Vuvan, & Alain, 2018). It has been shown that the thresholds of detection for pure tones and vowels in noise and are improved by explicit attentional cues and implicit contextual cues (Wolmetz & Elhilali, 2016). There is also a phenomenon called attentional blindness, where the perception of a target sound impedes the detection of another sound close to the target sound when auditory stimuli are presented in a rapid succession. This phenomenon is thought to be due to the interaction between higher processing centres associated with memory and attention (Shen et al., 2018).

Shafiro (2008) trained participants to recognise spectrally-degraded sounds and tested the accuracy of identification on novel sounds (not used in training) and the sounds used in training. It found that the accuracy improved for both types of sounds, indicating that the learnt skills (implicit memory) influenced accuracy of identifying brief sounds. This may be due to perceptual adaptation where participants learn to recognise certain acoustic cues like envelope or spectral cues properties and associate them with certain precepts of sounds (Shafiro, 2008; Gygi, Kidd, & Watson, 2004).

### 1.9 Previous studies on temporal resolution of short sounds

There have been multiple studies that have tested the time required to identify, detect, or classify short sounds using various methods for various purposes. The studies have either used a single sound class for identification of sound, or multiple sound classes for classification of sound types. There have also been numerous studies that have developed incredibly fast software/algorithms for identification and classification of very brief sounds (Lavner & Ruinskiy, 2009; Krishnamoorthy & Kumar, 2011; Yook et al., 2015).

Ballas (1993) tested the time required to correctly identify environmental sounds which it defined as the sounds or noises from non-living things, excluding musical instrument. There were 41 sounds in total, like door bell ringing, car horn, cooking sounds, water sounds among others. The participants were asked to identify the source of sound as fast as possible. The participants pressed a button to start listening to the sounds and pressed it again to stop the sound, when they had identified it. The time between the first and the second button press was called the identification time. After identification, each person had to type the source of the sound. They found that the shortest time taken to correctly identify was for the sound of ringing telephone at 1253 ms. The longest time was for the sound of an electronic lock at 6823 ms. The reaction time required to initiate and perform movement for button press, after the mental identification of sound, was added within the identification time. Therefore, the identification time also includes the reaction time of the motor processing and motor actions for hand movements which varies across individuals (Klemmer, 1956). Thus, for a more accurate identification time, a correction factor for motor reaction time needs to be added in or a procedure that excludes it must be used. The former might be much more unrealistic compared to the latter, as the motor reaction time varies across individuals and tasks (Klemmer, 1956).

Thiesen, Kopiez, Reuter, Czedik-Eysenberg, and Schlemmer (2016) reported data on duration thresholds for classification of music genre. Individuals could classify music sounds to their correct genre even at a short duration of 250 ms. It has also been shown that the participants were more accurate at classifying stimuli when only the instrumental part of a song were presented compared to stimuli where both vocal and instruments were presented (Gjerdigen & Perrot, 2008). Individuals without hearing problems could discriminate the emotional nature of classical music at very short duration ranging from 0.5 to 7.5 seconds. Even at the shortest duration, they were able to discriminate them confidently (Peretz, Gagnon, & Bouchard, 1998). Participants with normal hearing could identify the artist and the title of the songs presented in 400 ms long sound clips. Reducing the duration to 300 ms reduced the accuracy of identification of title and the artists, but some participants still identified the songs at above chance level (Krumhansl, 2010).

Gray (1942), presented vowel sounds of short durations to a group of participants. It showed that some vowel sounds could be recognised by some individuals at duration of as little as 3 ms. Moradi, Lidestam, and Rönnberg (2013) used gated stimuli in which they increased the duration of stimuli by 16.67 ms until the individuals correctly identified the consonants. The consonants were presented between two identical vowels. For example, if the stimuli was “aVa”, participants had to correctly identify the consonant ‘V’. The duration required to identify the consonants in quiet was significantly shorter that required in noise. However, in the beginning of the stimuli there was always a vowel sound, therefore the true duration of exposure to the consonant would be less than the actual threshold for their detection.

Ogg, et al. (2017) found that the participants were able to classify music, speech, and environmental sounds to their groups even when the sounds they heard were just 25ms long. The sounds used were tested using gated go/no -go paradigm, starting with 12.5ms duration and doubled in duration for each “gates”. The go/no-go paradigm requires the participants to respond only when the target sound is played and not respond when other sounds are played (Gomez, Ratcliff, & Perea, 2007). The minimum duration of the stimuli used was 12.5 ms and it did not test the ability to classify sounds below 12.5 ms. Therefore, the celling effect could be easily reached, as the speech sounds have been reportedly identified well below 5ms (Gray, 1942). Studies that use gated procedure only estimate the time range between which the sounds are identified, which means that there is always a risk of overestimating the identification thresholds. For example, in the study of Ogg et al. (2017), if a time threshold for identification was 15ms, the participant does not respond at gated stimuli of 12.5ms but they do respond at other stimuli at 25ms. However, the true threshold is in between those two gated stimuli.

A pilot study by Obert and Tchorz (2018) used a different procedure to find the thresholds for the individuals who reported to have normal hearing. The procedure used in this research also excluded motor reaction time when the time required for classification was calculated. Participants were required to listen to short sounds in a quiet environment and classify them into 4 categories: speech, noise, animal, and music. It used a four alternative forced choice method (4AFC). There were 40 trials for each sound class, totalling 160 sound files. The duration of sound started at 500ms and linear weighted up-and-down procedure was used. In the subsequent trials, the duration of the presentation of sound was reduced by 50ms if the responses were right and increased by 50ms if they were wrong. The 50ms step size was used until the duration of sound presented was 200ms. The step size was decreased to 10ms for presentation time shorter than 200ms. The up-and-down procedure for each sound class was independent of each other and the duration threshold was also calculated independently by averaging the durations of last 10 sounds for each class. The average time for classification of speech sounds was 21ms, for noise it was 24ms, 45ms and 55ms for music and animal respectively. However, the research encountered a celling effect as the step size was 10ms. This meant that the participants performance could have been better than 10ms in some trials, but the protocol could not present sounds below the duration of 10ms. The research relied on self-reporting to determine hearing ability. In addition, the speech sounds presented were in English whereas all the participants were German university students.

### 1.10 Aim and research questions

As discussed above, individuals with sensorineural hearing loss have poorer temporal resolution as indicated by poorer ability in gap detection and masking tests. Their temporal resolution is especially poor in rapid fluctuating noise. This inability might cause them difficulty in following the temporal architecture of the sound leading to decreased understanding of speech, identifying music and environmental sounds (Moore, 2008). Poorer temporal resolution in gap detection has also been shown to correlate with speech perception in noise (Gordon-Salant, & Fitzgibbons, 1993). Although hearing loss hinders temporal resolution, it is poorly predicted by pure tone thresholds (Glasberg et al., 1987).

In addition, frequency discrimination, which plays a major role in perception of pitch as well as speech perception, is impaired in individuals with hearing loss. Moreover, frequency selectivity is also impaired in those with hearing loss. Sub-normal frequency selectivity impairs perceptual abilities resulting in greater masking of a signal by background sounds. This in part impairs the ability of people with SNHL in understanding speech in noise as well as in the perception of timbre of music sounds (Moore, 2008).

Most of the studies that individually test temporal resolution, frequency discrimination and frequency selectivity used a stimulus with few characteristics like pure tones and narrow band noise and test the ability to detect (Snell et al., 2002; Tyler et al., 1982). In contrast, this study wanted to test auditory temporal resolution using short sounds which contained complex information and need more complex auditory processing skills.

Multiple studies have tested the time required to identify, detect, classify short sounds using various methods as mentioned in section 1.9. However, the durations thresholds which are needed to identify different sound types, such as music, speech, noise, and animal sounds in normal hearing and HI subject have not been investigated. Their relation to speech understanding in noise, and hearing ability are also not clear. Knowing the temporal resolution for short sounds will allow us to investigate what the minimum psychoacoustical properties are needed for the recognition of different types of sound. It would also allow us to investigate whether the ability to recognise sounds in short duration helps with the ability to understand speech in noise.

This study aimed to investigate this by building on the research by Obert and Tchorz (2018), and by exploring the procedure that can reliably measure the minimum time needed to classify the four major categories of sounds (will be referred as classification time thresholds from here on). In addition, this study aimed to investigate whether the classification time thresholds correlates with pure tone average (PTA) thresholds and SRT. The following main research questions were posed in order to investigate our aims:

- What time durations are the thresholds for classifying speech, music, animal, and noise, respectively for hearing impaired and normal hearing individuals?
- Are classification time thresholds correlated to pure-tone thresholds?
- Are classification time thresholds correlated to SRT?
- Is SRT correlated to pure-tone thresholds?

For the questions above, following null hypotheses were formulated and statistically tested with alpha level of $p = 0.05$:

**$H_0$1:** There will be no significant differences between the classification time thresholds for participants with hearing impairment and participants with normal hearing for speech, noise, music, and animal sounds.

**$H_0$2:** The classification time thresholds will not be correlated to pure tone thresholds.

**$H_0$3:** There will be no correlation between SRT and PTA thresholds.

**$H_0$4:** The classification time thresholds will not be correlated to SRT thresholds.

**$H_0$5:** There will be no correlation between Age and PTA thresholds.

**$H_0$6:** The classification time thresholds will not be correlated to age.

**$H_0$7:** There will be no significant difference between Run A and Run B of classification time thresholds.

**$H_0$8:** There will be no significant difference between classification time thresholds with feedback and classification time thresholds without feedback.

# 2. Methods

## 2.1 Overview

The aim of this project was to investigate the minimum time it takes for normal hearing (NH) and hearing impaired (HI) participants to identify sound as belonging to one of 4 different categories: speech, noise, music, and animal. This project extended and improved on the project by Obert and Tchorz (2018). It focused partly on bettering the procedure for determining classification time thresholds and partly on investigating the relationship between these thresholds and pure tone average (PTA) thresholds, and speech reception in noise thresholds (SRT). Part A and Part B are focused on developing the procedure for measurement of classification time thresholds. Whereas, the data gathered from the procedures used in Part C will be used to answer the questions posed in Section 1.10.

## 2.2 Part A: Pilot study for determining classification time thresholds

The purpose of pilot study was to develop a modified procedure based on Obert and Tchorz (2018) and test the procedure in few normal hearing participants to determine whether it would produce reliable results and troubleshoot any problems that are encountered before testing in a larger group of participants.

### 2.2.1 Participants

The participants for the pilot study were Audiology students from University of Canterbury recruited via an online group forum with incentive of 10 dollars petrol vouchers. There were 3 participants, 2 male and a female with the age range of 22 to 27 years old. All participants were native New Zealand speakers, had normal hearing, and good dexterity and eyesight to operate touch screen monitor.

### 2.2.3 Stimuli

Several changes were made to the procedure of Obert and Tchorz (2018), but few were also kept the same for the purposes of comparison. The audio files, for this part of the research, were the same as those used by the Obert and Tchorz (2018). There were 160 sound files in total: 40 files for each sound class. The software shuffled 16 sound files, 4 from each category, from the sound bank which were then presented in a random order and this process was repeated 10 times. The presentation level for the sound was 65 dB A, and the same signal was presented binaurally. For each sound class there was one staircase, resulting in four simultaneous running staircases. The sound files were cropped from the front end to present them in an appropriated duration. A sound level equalisation, specific to the headphones being used in the study, was done in LabVIEW (National Instruments, TX, U.S.A) on the long uncut sound files. Due to the difficulties in measuring the sound level of audio samples that were so brief, the sound level of the uncropped file was adjusted to 65 dB A prior to cropping. To avoid audible spectral splatter at the onset and offset of the soundfiles, ramps of 2.5 ms were applied to the beginning and the end of each cropped audio file.

There was no strict definition for the classes of sound in this version. However, speech had only one speaker either male or female speaking clearly without any background noise. Music consisted of many genres including both vocals, vocals, and instrumentals, and purely instrumentals sounds. Animal noise included one animal noise at a time. It included mammal, birds, and insects. Noise was any disorganised sound that could not fit into one of the categories above, and the category included, synthetic and natural noises. As another change in procedure, any sound file that sounded ambiguous (e.g. tonal animal calls that could be perceived as music) were removed.

The adaptive procedure for threshold seeking was changed. In the previous study, a linear method was used to change the step size for the duration of sound presented. Obert and Tchorz, (2018) encountered ceiling effects, as when the duration approached the same size as the step, it was not possible to decrease the duration any further. This limited the accurate determination of the performance of the participants, as the adaptive procedure converged on a duration that enabled the participants to give a better performance.

In this research, an adaptive weighted up/down staircase method adapted from Kaernbach (1991) was used. Due to the use of 4-AFC, the midpoint of the psychometric function was 62.5% as the minimum score by chance was 25% and maximum possible score is 100%. The equation, Sdown p = $S_{up}$ (1-p), where $S_{down}$ is step down, $S_{up}$ is step up and 'p' is the point of convergence, was used.

When the participant responded correctly, the presented duration of the sound presented was multiplied by $S_{down}$, so that the duration of sound for the next trial became shorter. In contrast, when the response was wrong the duration of the sound was multiplied by $S_{up}$ so the duration of sound for the next trial became longer. There were two different phases of the run, in which two different step sizes were used. In the initial phase, step sizes were bigger to approach the threshold quicker. Whereas, in the working phase, the step sizes were smaller to improve the accuracy by having the durations fluctuate around the true threshold for classification. For each sound class, the initial phase lasted from the first trial, until the 15$^{th}$ trials, where the $S_{down}$ was 0.75, and the $S_{up}$ was 1.42. The working phase lasted from trial 16$^{th}$ till the 40$^{th}$ trial, where the $S_{down}$ was 0.875 and $S_{up}$ was 1.2085. The thresholds for each sound classes were obtained by averaging reversals, in the last 20 trials of the working phase as shown in the Figure 2.1. The initial phase, from trial 1 to 20, has larger steps and helps get to the threshold quicker. Whereas, the working phase, from trial 21 to 40, has smaller step sizes with the durations of presentation fluctuates around the true threshold (solid blue line) providing more accurate measurement. The thresholds are measured by averaging reversals as shown by the black arrows from the working phase. A reversal is defined as a point where the response goes from being right to wrong (upward arrow), or from being wrong to right (down ward arrow).

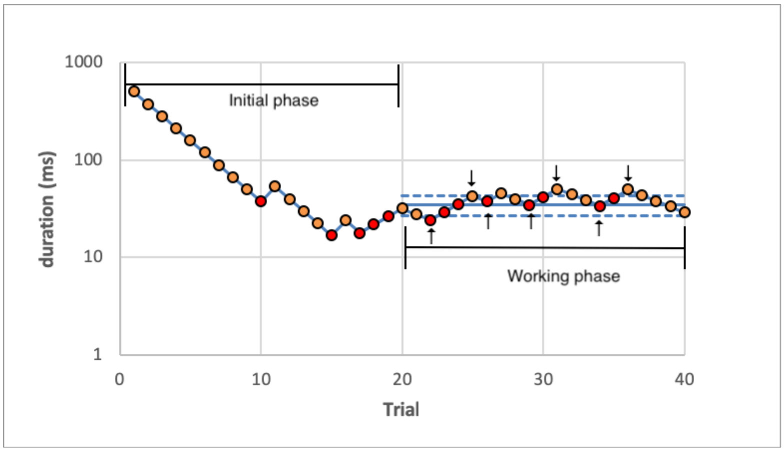



*Figure 2.1.* An example of the up and down procedure. The slope descends when responses are right and ascends when wrong.

### 2.2.4 Materials

The classification time threshold experiments were carried out software custom-written using LabVIEW, (National Instruments, TX, USA). A Sound Blaster X-Fi Surround 5.1 Pro external sound card (Creative Labs, Singapore) was used with Sennheiser HD 215 headphones (Sennheiser electronic GmbH & Co). The software was installed on HP Elite 800 G1 TWR running on Windows 7, connected to two monitors. A touch screen touch monitor (ELO ET1715L; Tyco Electronics Corp., USA), was located inside the booth with the participants and was used to display the response options and the record the response.

### 2.2.5 Procedure.

Each participant was tested on their own in a sound treated room. They were given verbal instruction, and written instructions shown on the touchscreen display monitor. See Figure 2.2 for the instructions displayed on the screen before the testing. They were instructed that they would hear 160 short sounds individually consisting of four different types; 40 each of speech, noise, music, and animal. For each sound, they were to choose which one of the four categories the audio clip sounded like and respond by touching one of the four options on the screen. After they choose one, they had brief time to change it, but if that time passed their response was locked in. After their final response was accepted, they received visual feedback: tick for right and cross for wrong. The images of the feedback icons are shown in Figure 2.3. They were also told that after the feedback disappears, they would hear another sound. The audio clips would gradually get shorter and shorter. They were to raise their hand if they did not hear any sounds or had any questions. The researcher supervised the participants from outside the sound booth, through a glass window, and could hear any queries from the participants. The participants repeated the procedure twice, each taking around 15 minutes. After tests were finished, they were asked what the strategies they used for categorising the sounds.

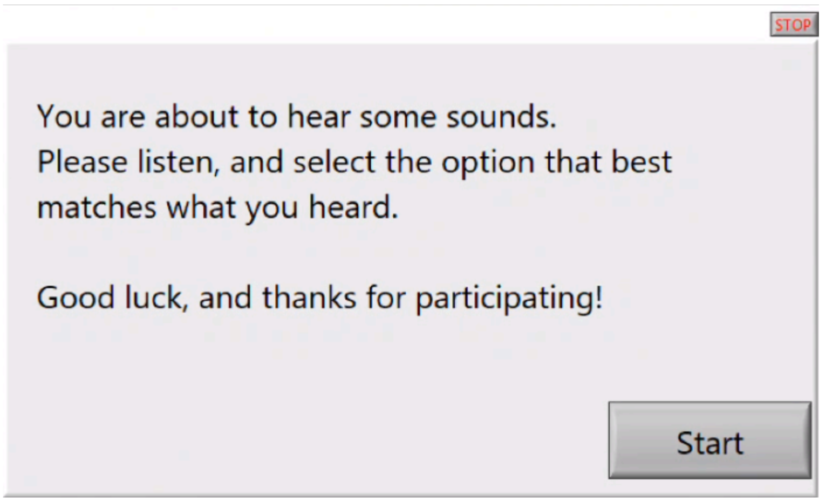



*Figure 2.2.* Screenshot of the information shown to the participant before the testing started.

*Figure 2.3.* The feedback icons displayed after correct response (green tick) and after wrong response (red cross).

**2.2.6 Problems encountered in the pilot study.**

Prior to the pilot study, the sound files were checked individually by the researchers. It was found that the quality of sound for the four different types of sound was vastly different. The speech sounds were all taken from audiobooks or interviews recorded in a studio and were of good quality, with a wide bandwidth. Music sounds were diverse, taken from a range of different genres. The animal sounds were of very poor quality, with low bandwidth, and often included some degree of background noise as most of them were recorded outdoors. The subjective reports from the participants confirmed that by attending solely to the recording quality and bandwidth, the participants could often successfully choose the correct category of the sounds without reference to the actual content of the sound files. The participants reported that the noise sounds were easiest to identify, as everything that sounded like white noise/static was considered noise. Also, if they were not sure of the category, many would choose noise as by default. They reported that the music was anything that sounded sharper (i.e. contained higher frequencies). Speech was easy to identify because the studio recordings were equalized in a way that emphasised the lower bass frequencies. Animal sounds were reported as the most difficult to classify.

The participants also reported that some sounds got very short around the end, they only heard brief clicks but could still successfully identify which category it belonged to. When the data was examined, it was found that some adaptive tracks lacked any reversals at all, and in some sound class the participants were recognising the sound category nearly 100% of the time, even when the sounds were below 5 ms. This was unrealistic as the ramp was 5 ms in total: 2.5 ms at the beginning and 2.5 ms at the end of the sound. Therefore, it is difficult to argue that the thresholds obtained from the traces with an unrealistically high number of correct responses are in anyway representative of the participant's true categorisation threshold. Figure 2.4 shows an example of such a trace. The noise sounds are identified 95% of the time and only two mistakes (indicated by red data points) are made in the initial phase. The multiple correct responses are found well below 5ms duration which is the duration of the ramps. The bar graph shows the percentage of response made for each sound class.

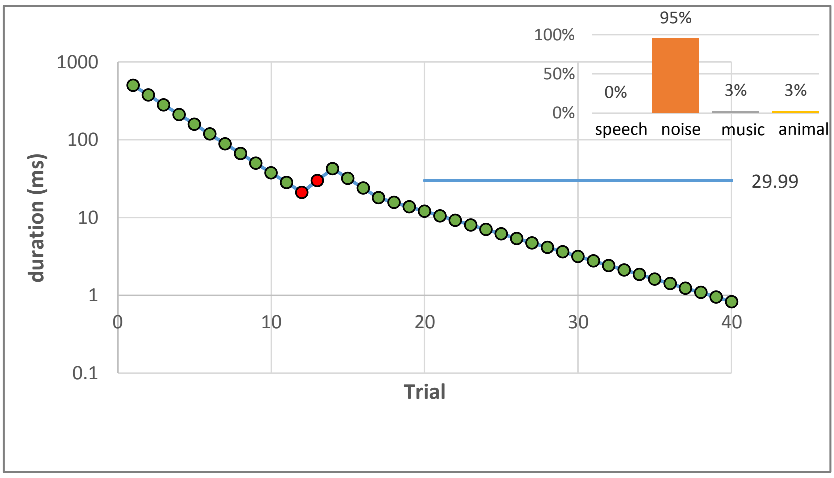



*Figure 2.4.* An example of raw data where the number of correct responses exceeds realistic expectation.

As there are four simultaneous staircases, once one particular sound class differed in duration from others sound class by a certain amount, it would be easy for the participants to recognise that sound class based on its duration instead of its content. In addition, they received correct feedback when they identified short sound as being that one particular class, providing a distinct advantage. It seemed that the feedback was perpetuating the problem by acting as reinforcer for shorter sound belonging to a particular class. This advantage only grows stronger as the trials proceeds because with every right response for that particular class, the sound gets shorter and it becomes easier to distinguish from other sound classes. This trend can be clearly be seen in Figure 2.5. The noise sounds get distinctively shorter than other sound classes, the duration differences from other classes only grows larger as the trial proceeds. As the trial proceeds the difference in duration between noise and other sound classes grows which makes it easier to identify noise sounds based on durational cue, resulting in unrealistic number of correct responses and lack of reversals in the working phase.

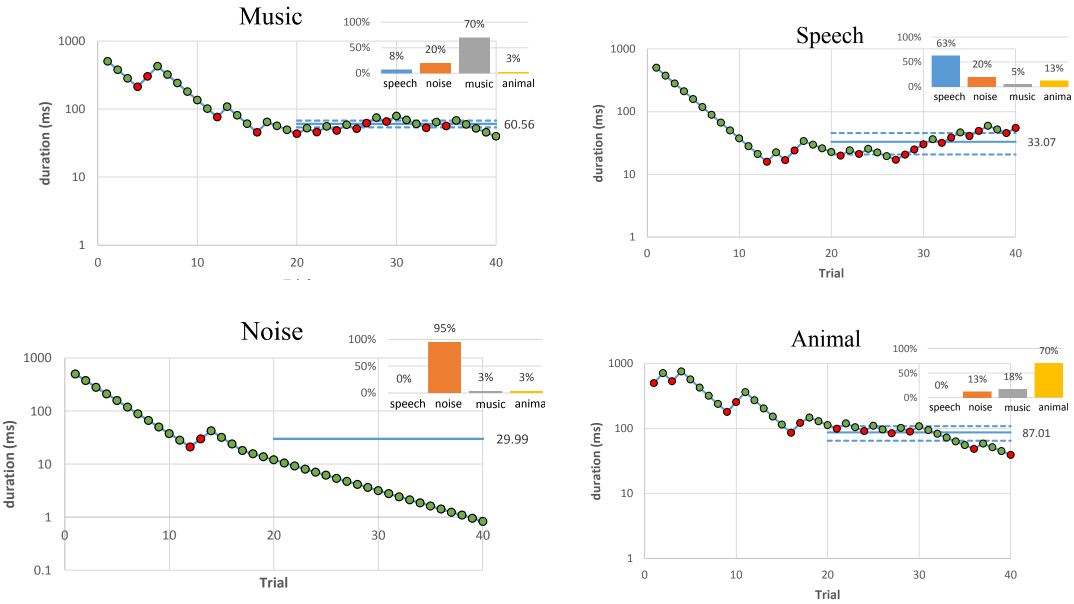



*Figure 2.5.* An example of staircases where there is a separation in duration as the run proceeds.

In summary, the two main problems encountered were i) poor sound quality that resulted in participants classifying sounds based on the acoustic properties of the sound recording rather than the content of the sounds themselves; and ii) that some of the participants were achieving 100% correct responses for some sound classes, particularly noise, resulting in lack of reversals. This was thought to be due to one class of sound getting ahead of others, in the 4 simultaneously running staircases.

## 2.3 Part B: Attempts to solve the problems encountered in pilot study for time classification thresholds.

### 2.3.1 Overview

This part of the study attempted to solve the problems encountered in the pilot study. Significant changes were made to stimuli, while procedure and the materials were kept constant. The same testing site, and instructions used in Part A were used. The number of participants were increased and included both hearing impaired and normal hearing.

### 2.3.1 Participants

Participants with normal hearing loss were recruited via a group forum, where the information sheet and invitation letters were posted. Ten participants responded to the forum and booked to participate in the study. Participants with hearing loss were recruited via University of Canterbury, Speech and Hearing Clinic (UCSHC). The contact details of the potential participants who had already agreed to be contacted for research were taken from the volunteer's file in UCSHC. The proposal to do so was mentioned in the ethics application to the University of Canterbury Human Ethics Committee (see Appendix A).

The clinic file had recent audiometric information along with their contact details. Only the details of clients who had consented to be contacted, who were native New Zealand English speakers, and were aged above 18 years old were invited to participate. They also had to fulfil following criteria:

1. Pure tone thresholds asymmetry of no more than 20 dB at 2 different frequencies, or 30 dB at any single frequency
2. No conductive component of 15 dB or more at more than two different frequencies, or 20 dB or more at one frequency across both ears.

From the volunteer's folder, 40 potential participants were identified in the manual search, and their names, addresses, and emails if they had provided were recorded. Nineteen of the 40 participants had supplied email addresses and were contacted via email in which they were sent an invitation letter, information sheet, and consent sheet, along with a brief explanation for the purpose of the email. See Appendix A for a copy of the invitation letter information sheet, consent sheet. Only 1 person replied to the email and booked for testing. Twenty-one out of 40 people were contacted via letter. An invitation letter information sheet, and a small sheet explaining the ways to contact the researchers, were sent to the invitees along with a paid return envelope. Six people out of 21 responded and were subsequently booked for the testing and attended the appointment. Two persons were recruited by word of mouth.

Before the invitees were booked, they were asked if they had memory issues, significant eye sight loss, or dexterity issues, discharging/infectious ears, recent ear/head surgery. None of the contacted person reported the issues.

In total, there were 19 participants, 10 of them were female and 9 were male, with an average age of 45.3 years, ranging from 22.3 to 84.9 years old. Ten of them had normal hearing and were from within the University of Canterbury, and 9 of them had hearing loss and were recruited from outside the university.

### 2.3.2 Stimuli

To solve the problem encountered in pilot study, where participants seemed to classify the audio files based on quality of sound, it was decided that all the sound files should be replaced with new ones that were diverse in quality in each category. Strict criteria were used to choose the files. A bank of sounds was obtained from various sources. Noise sounds were mostly obtained from online archive of British Broadcasting Corporation (BBC, 2018). Animal sounds were obtained from YouTube, BBC sound effects, as well as the digital animal sound archive of the Museum für Naturkunde, Berlin (Tierstimmenarchiv, 2018). All the music was obtained from YouTube. Each class of sounds were sub-categorised arbitrarily in an attempted to diversify the collection of the sounds. Most of the sounds aimed to be familiar to the participants and reflect sounds of the everyday experience.

For music, attempts were made to capture as many genres as possible. Half of the music files were instrumental, and the other half were non-instrumental. The speech files were also diverse, it included sounds recorded by different types of devices and of different qualities. The speakers included school children, news persons and interviews. Half of the speakers were male, and the other half were female. All the speech sounds had no background noise and only had a single speaker speaking in decipherable English. Most of the speakers had New Zealand accent but a few also had foreign accent.

The animal sounds were also chosen to represent diversity, in quality and frequency spectrum. There were birds, pets, wild and domesticated animals, and even insects. However, a balance between diversification and familiarity was hard to find for the animal sounds. Noise was also diversified and included several varieties of sounds. The noisy sounds are usually dominated by low frequency sounds (Wagener & Brand, 2005). It was made sure that there were high frequency sounds included in noises like glass breaking, paper noises, high frequencies radio noise, and running water. Refer to Appendix B to see the complete list of sounds used.

Once the sounds were converted into compatible format, they were edited using Audacity (Audacity(R), General Public License). They were all trimmed to be between 500 ms and 1000 ms long. Any sounds containing any significant temporal gaps were removed. For example, slow footsteps, or a significant pause in speech between words. The bandwidths of each sounds were checked, and it was made sure that it was more than 10 kHz with no frequency clipping, to ensure certain standard of quality. The sounds were also checked for amplitude clipping, any sounds with amplitude clipping were removed. Example of sounds excluded due to frequency limitations and peak clipping is shown in Figure 2.6 and Figure 2.7 respectively.

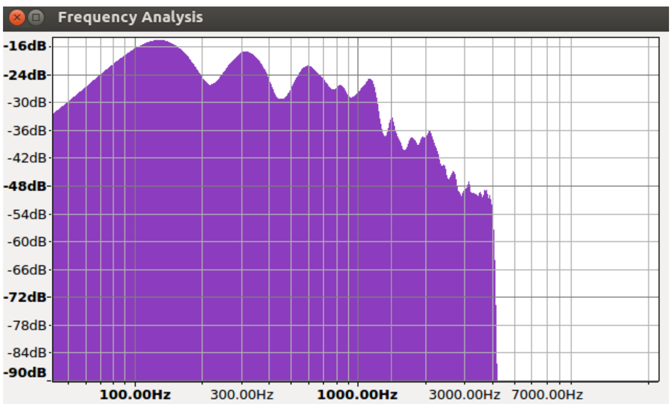



*Figure 2.6.* A screenshot of frequency analysis from Audacity showing an upper frequency bandwidth limit at 4000 Hz indicated by sharp drop off in the spectrum.

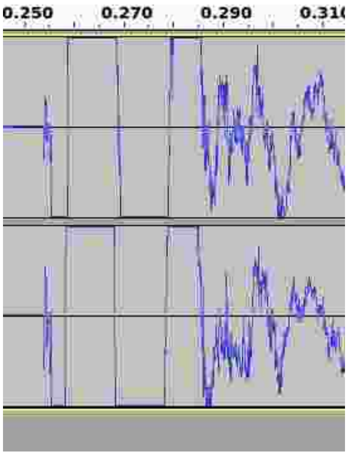



*Figure 2.7.* A screenshot of amplitude analysis from Audacity, for gunshot noise, showing peak clipping between around 0.260 ms and 0.285 ms.

The sound files that sounded ambiguous were also removed. Sounds from one class, that could also be subjectively perceived like other class were considered ambiguous. For example, in the process of removing ambiguous sounds, a wolf howling that might sound like a siren from a distance was removed, among others. The sounds files were checked and rechecked and approved by two other persons involved in the research. There were 40 sounds files in each class, totalling 160 for the test trials.

To address the issue described above whereby one staircase gets away from the others, becoming noticeably shorter and providing a durational advantage, we changed three things. Firstly, we decreased the number of files that the software randomises at a time from 16 to 8. In the pilot study, the software took 16 random files from the bank – 4 sounds from each of the 4 categories – and presented them in random order. This was preferable to choose the files completely at random, but even with these 16 files the participant could have had 4 subsequent presentations of same class of sound. If this happened for any one of the sound class, the presentation durations of that class would decrease faster than all the other classes significantly. This was more problematic at the beginning where the step sizes are larger. Once the duration for that class separated from the others, it would allow a durational advantage. To prevent the chances of a high number of subsequent presentations of same sound class, we decreased the number of files randomised from 16 to 8. This meant that only 8 files were randomised at once. Two files from each of 4 sound classes are taken and randomised for presentation, decreasing the chances of one particular class getting ahead of others based on increased number of subsequent presentations of the same class.

Secondly, to prevent them from using durational cues to classify sounds, we added 20% extra audio files, as decoy trials. The decoys had the same characteristics as the trial sound files but did not form part of the 160 files to be presented. The decoys were always played at the same duration as the duration of class with the longest duration at that time. The responses for the decoys were not recorded when calculating classification time thresholds. The rationale for adding in decoys was that if a participant were listening to the class with the shortest duration and it was getting shorter with every presentation, at least 1 in 5 presentation of that sound class would have same duration one of the other class, therefore, taking away some of the durational cues.

Thirdly, we removed feedback to prevent a positive feedback loop from forming with the sample duration. The rationale for this was that the feedback was providing enforcement that the sounds with shorter duration were of one particular class, as described above. It was thought that, if the feedback was removed it could remove that reinforcement and prevent the feedback loop.

### 2.3.3 Equipment and Procedures

The equipment was kept constant to the pilot study. Please refer to section 2.2.4 for details. Each participant was tested on their own in a sound treated room. They were given verbal instructions and in written form in the touchscreen display monitor. They were instructed that they would hear 200 short sounds individually. There would be four different types of sound 50 each of speech, noise, music, and animal. They were to choose what the sound they heard closely resembled by touching one of the four options on the screen. There was no time limit within which they had to decide after the presentation of a stimulus. After they choose one of the options, they would still have a very brief time to change the response by choosing one of the other options, if that time passed, they would not be able to change response and they would hear the next sound. They were also told that the sound would gradually get shorter and shorter. They were instructed to stop and report if they did not hear any sounds or had any questions. If the client reported they were not hearing sounds or were missing a significant number of sounds, the presentation level was increased to either 70 dB or 75 dB and the test was repeated from the start. The researcher supervised the participants from outside the sound booth with the glass window and could hear any queries from the participants though speakers. The participants repeated the procedure twice, each taking around 15-20 minutes each.

**2.3.4 The problems encountered in Part B.**

There was an unintended consequence of removing feedback: participants tended to choose one option over others by default. Whenever they were uncertain, they tended to choose noise. This resulted in the noise sounds becoming shorter and separated from others. The test was then no longer a four alternative forced choice, as they were always more likely to choose one particular sound class. This can be seen in the Figure 2.8 – the percentage of wrong responses was the highest for noise in all the three other classes of sound. The highest proportion of mistakes in other sound classes was committed by choosing noise. To fix this problem, feedback was added back to the experimental procedure. The rationale for it was that the negative feedback might discourage people from constantly choosing one particular class when they were unsure. The risk of a forming a feedback loop was accepted and it was decided that those durational thresholds where there were lack of reversals in the last 20 trials, or where percentage correct reached more than 80% at the last 20 trials would be removed.

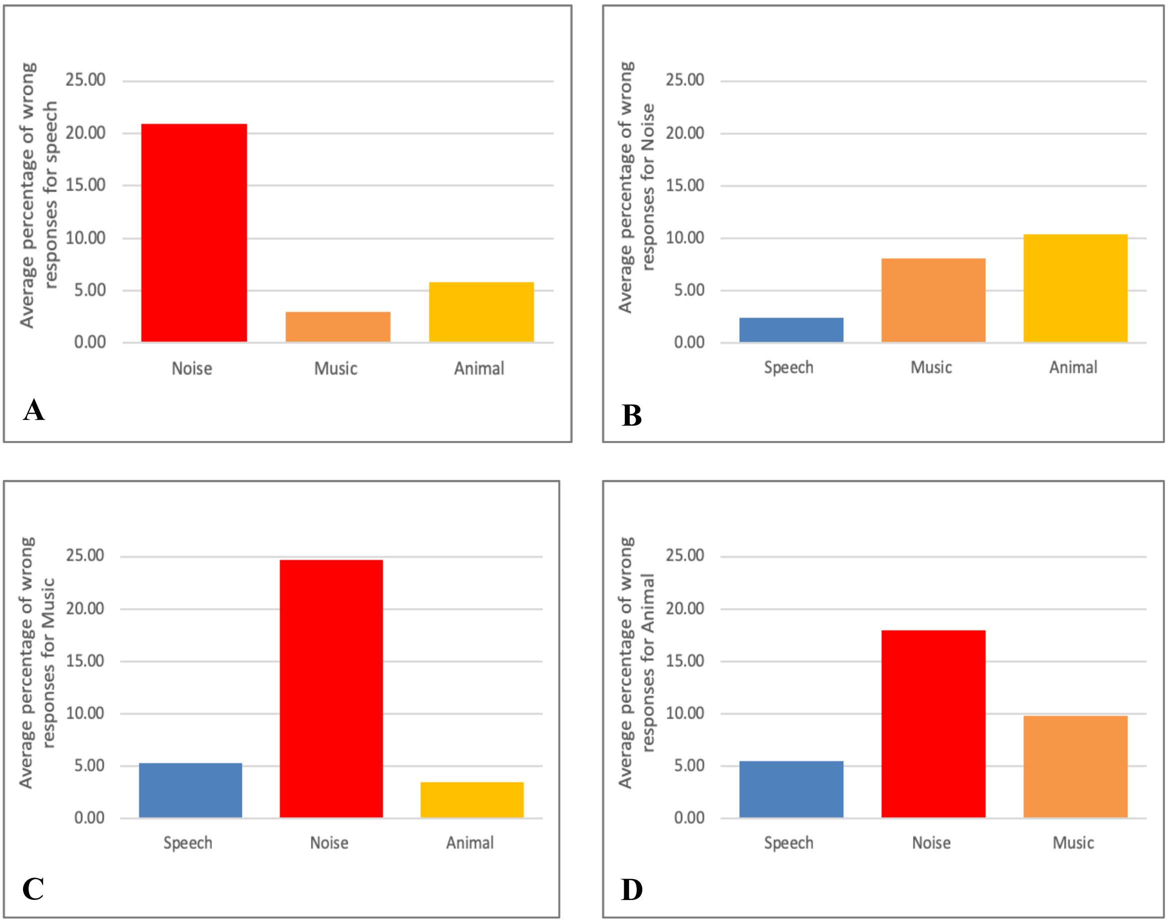



*Figure 2.8.* The percentage of wrong responses made for other sound classes. As seen in A, C, and D, the percentage of wrong response made by choosing noise is very high in all other sound classes. (A. Mistakes made for speech. B. Mistakes made for Noise. C. Mistakes made for Music. D. Mistakes made for Animal).

## 2.4 Part C: Final protocol for classification time thresholds and additional test; PTA and speech in noise test.

### 2.4.1 Overview

In this part of the study, a final protocol was formed for testing classification time thresholds. The stimuli, materials and procedure for this part was kept constant to part B except that the feedback was added back on. However, significant changes were made compared to the Obert and Tchorz (2018), which are described below. More participants were recruited compared to Part B. Two more tests were administered: New Zealand speech in noise matrix test and audiometric pure tone threshold testing. The first administered was audiometric test, followed by first run of classification time threshold measurement. This was followed by Speech in noise test. Lastly, the second run of classification time threshold was tested. The data from Part C were used to answer the questions posed in Section 1.10.

### 2.4.2 Participants

There were 44 participants in total, 16 of them had normal hearing and 28 of them had hearing loss. HearForm (HearForm Software, LLC, Northport, Washington, USA) was used to search for clients, it had their recent audiograms, the age, and whether the clients gave consent to be contacted for research. Only those who had consented to be contacted were listed to be contacted. Clients with asymmetrical hearing loss, of 20 dB or more at 2 difference frequencies, or 30 dB or more at single frequency were ignored, as well as those with conductive loss of more than 15 dB at two or more frequencies, or more than 20 dB at any one frequency across both ears. To be listed to be contacted the clients had to be native New Zealand English speaker aged above 18 years old.

The primary contact details of 73 potential clients were listed with different degrees of hearing loss. Sixty-six of them supplied their phone number and 7 had listed only their email. Out of 7 emailed, with the information sheet and invitation letter, along with short expiation for the purpose of the email, only 2 replied and 1 was interested and booked for a testing and the other had moved. Those who had supplied phone number, were contacted during work hours, the nature of the research was described over the phone and all their questions, if they had any, were answered before they were booked in for a testing appointment. Out of 66 clients phoned, 20 of them were interested in the research and booked for an appointment.

Six students from within the university participated in the research – 4 of them had also participated in the previous pilot, and 2 of the new participants were among those recruited via word of mouth. In total, thirteen participants were recruited via word of mouth. Most people who were recruited via word of mouth had normal hearing. Eight out of nine of participants with hearing loss who had participated in the previous pilot also responded for the second set of measurements. Those that participated again were also reimbursed with the same amount again.

In total, there were 46 people booked in for the appointment, 44 attended the appointment. Refer to Table 1 for summary. Sixteen of them had normal hearing and 28 of them had hearing loss. Twenty-one of them were male and 23 were female. The mean age was 55.0 years old and ranged between 21.1 and 89.0 years old.

**Table 1.**

*Summary of ways that participants were contacted and recruited.*

| Ways of contact | Number of people contacted | Number of people Recruited | Number of participants attended |
|---|---|---|---|
| Phone | 66 | 20 | 18 |
| Email | 7 | 1 | 1 |
| Previous Research | 13 | 12 | 12 |
| Word of mouth | 13 | 12 | 13 |
| Total | 99 | 46 | 44 |

**2.4.3 Audiometric assessment.**

The data was collected in the double walled, sound treated booth in UCSHC, Christchurch, New Zealand. A GSI-61 diagnostic audiometer (Grason-Stadler Corp., USA) was used for threshold collections. ER-3A insert earphone was used as a transducer for obtaining air conduction thresholds, whereas Radioear B-71 bone conduction vibrator was used for measurement of bone conduction thresholds. The thresholds were collected using standard New Zealand Audiological Society protocols (New Zealand Audiological Society, 2007), and tested 250 Hz, 500 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz, in each ear individually. Air conduction (AC) and bone conduction (BC) thresholds were measured using modified Hughson-Westlake (Carhart & Jerger, 1959). BC thresholds was only measured in

500 Hz, 1 kHz, 2 kHz, 4 kHz where with AC thresholds were above or at 20 dB HL. In this study PTA thresholds was calculated by averaging of thresholds at 500 Hz, 1 kHz, 2 kHz, and 3 kHz. High frequency PTA (HFPTA) was calculated by averaging 4 kHz, 6 kHz, and 8 kHz. Hearing loss was defined as having PTA or HFPTA of above 15 dB HL.

Participants with conductive loss of 15 dB at two or more different frequencies, or more than 20 dB at any one of the frequencies across both ears were excluded. Participants with asymmetry more than 20 dB at two frequencies, or more than 30 dB at a single frequency were also excluded.

### 2.4.4 Speech in Noise

Speech in noise test was performed using the University of Canterbury Auditory-Visual Matrix Sentence Test (UCAMST) in auditory-alone mode (Trounson, 2012). A Sound Blaster X-Fi Surround 5.1 Pro (Creative Labs, Singapore) external soundcard was used, along with Sennheiser HD 215 headphones (Sennheiser Electronic GmbH & Co, Wedemark, Germany). UCAMST was installed on HP Elite 800 G1 TWR (Hewlett Packard Enterprise, California, United States) running on Windows 7 (Microsoft Corporation, Washington, United States).

Each sentence had five words with following structure: name, verb. number, adjective, and object in that order. For example, “Thomas likes three new books”. The sentences were synthesised from 50 words, 10 for each category in the structure, with possibility of forming 100,000 unique sentences which were semantically unpredictable and syntactically fixed. All sentences were grammatically correct and comprehensible. The speaker had New Zealand accent and artificial synthesis of sentences attempted to preserve natural coarticulations and prosody. The sentences were open set and the noise was constant. The individuals listened to 20 sentences with different signal to noise ratios that were adjusted via an adaptive procedure. The signal to noise ratio that was needed for them to understand 50% of the sentence was regarded as SRT in noise. This was also the midpoint of the psychometric curve.

The adaptive procedure used to find SRT was the A1 procedure developed by Brand and Kollmeier (2002), based on a generalization of the procedure of Hagerman and Kinnefors (1995). The procedure estimates the SRT by presenting sentences at two different SNRs. These are called "pair of compromise" (p1 = 0.19 and p2 = 0.81). The presentation level of subsequent trial is based on the responses by the participants for the trial. When the response is right the presentation level of the subsequent sentence is decreased and when the response is wrong presentation level is increased for the next trial. As the trials progresses, the presentation level converges towards the SRT of the participant. The final SRT is calculated by fitting a psychometric curve to the data.

The participants were instructed that they would hear five-word sentences with background noise which will fluctuate becoming louder and softer. They were to repeat whatever that they hear. They were told that they could struggle when the noise is loud, but if they cannot make a sentence, they were to just repeat the words that they hear. The sentences would all be sensible. The test took around 5-10 minutes.

### 2.4.5 Final protocol for classification time thresholds

#### 2.4.5.1 Stimuli

There were 200 files, 50 each for the four sound classes. Forty sounds from each sound class were used as real trials that counted towards calculation of time classification thresholds. Ten sound files from each class were used as decoys, which were not counted in the calculation for thresholds. The decoys were played once every fifth trial with the same duration as that of the class with longest running duration. As described in section 2.3.2, the sounds were obtained from various sources and it was made sure that each sound class had as many diverse sounds as possible to get a fair representation of that class. They were all trimmed to be between 500 ms and 1000 ms long using Audacity. Any sounds containing any significant temporal gaps were removed. The bandwidths of each sounds were checked, and it was made sure that it was more than 10 kHz with no frequency clipping, to ensure certain standard of quality. The presentation level for the sound was 65 dB A, and the same ramping and equalisation process was applied as was described above in Section 2.2.3. The adaptive weighted up/down staircase applied was the same as described in Section 2.2.3.

#### 2.4.5.2 Equipment and Procedures

The equipment was kept constant to the pilot study. The participants were tested in the same room along with the same equipment described in section 2.2.4. Same procedure used in section 2.3.3 was used for measurement of time classification thresholds was used except the feedback was added back to address the problem encountered in Part B, as described in section 2.3.4. They received: a green tick for correct and a red cross for wrong responses, as used for the pilot study. After the feedback disappeared, they would hear another sound clip. Another change was that they had a break in between Run A and Run B of testing, where a speech in noise matrix test was performed. Only changes to instructions were regarding the changes made above.

The time classification thresholds values that lacked reversals in the last 20 trials, or percentage correct reached more than 80% at the last 20 trials were excluded. The mean from two runs was calculated for each sound class to be used in the analysis. If one data value was excluded due to the reason mentioned above, just the remaining one was used.

### 2.4.6 Exclusion based on audiometric thresholds.

Four participants were excluded because of significant air-bone gap. Three participants were excluded because of significant asymmetry. One participant was excluded because of both air bone gap and significant asymmetry. Only one run of classification time thresholds was carried out for 6 participants due to time constraints

# 3. Results

## 3.1 Overview

The results consist of three parts. The data for Part A or the pilot study was not statistically analysed, but was used to recognise problems and troubleshoot methodical procedures, as mentioned in section 2.2.6. The data for Part B was analysed to determine the classification time thresholds for each sound class when tested without feedback. As mentioned in the Method section, the data extracted for classification time thresholds using feedback was used to answer the questions posed in the statement of problem. This was done in Part C, where it was analysed with data from pure tone average thresholds and speech reception in noise thresholds.

## 3.2 Part A: Pilot study

The pilot study only had 3 participants, therefore no statistical analysis was carried on them.

## 3.3 Part B: Classification time thresholds without feedback

In this section, the classification time thresholds that were measured without feedback are analysed. The data from the no feedback runs are also compared with the data from runs with feedback for those who participated in both tests. The main problem with measurement of classification time thresholds without feedback is also highlighted.

### 3.3.1 The classification time thresholds for runs without feedback

For this analysis, the mean values across the two runs were calculated for each participant. If a value from one run was excluded, the value from another run was used instead of the mean. Distribution of data for the four sound classification thresholds was checked as shown in Table 2. It was found that the data for speech, noise, and animal was normally distributed. Whereas the data for music was not normally distributed. Therefore, a non-parametric Wilcoxon Signed Ranks test was performed to investigate the performance of participants across the four classes of sound. The results are tabulated in Table 3.

**Table 2.**

*One-Sample Kolmogorov-Smirnov Test for time classification thresholds without feedback.*

| | Null Hypothesis | Significance level | Decision |
|---|---|---|---|
| $H_0 1$ | The distribution of nofeedback-speech is normal with mean 27.90 and standard deviation 10.540. | .068[1] | Retain null hypothesis |
| $H_0 2$ | The distribution of nofeedback-noise is normal with mean 60.66 and standard deviation 33.185. | .200[1, 2] | Retain null hypothesis |
| $H_0 3$ | The distribution of nofeedback-music with mean 61.36 and standard deviation 33.815. | 0.005[1] | Reject null hypothesis |
| $H_0 4$ | The distribution of nofeedback-animal is normal with mean 76.60 and standard deviation 52.516. | 0.200[1, 2] | Retain null hypothesis |

Asymptotic significances are displayed. The significance level is .05.
[1]Exact significance is displayed for this test.
[1]Lillefors Corrected.
[2]This is a lower bound of true significance.

**Table 3.**

*Wilcoxon signed ranks for classification time thresholds without feedback.*

| | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| nofeedbacknoise - nofeedbackspeech | Negative Ranks | 1 | 1.00 | 1.00 |
| | Positive Ranks | 5 | 4.00 | 20.00 |
| | Ties | 0 | | |
| | Total | 6 | | |
| nofeedbackmusic - nofeedbackspeech | Negative Ranks | 0 | .00 | .00 |
| | Positive Ranks | 14 | 7.50 | 105.00 |
| | Ties | 0 | | |
| | Total | 14 | | |
| nofeedbackanimal - nofeedbackspeech | Negative Ranks | 1 | 1.00 | 1.00 |
| | Positive Ranks | 11 | 7.00 | 77.00 |
| | Ties | 0 | | |
| | Total | 12 | | |
| nofeedbackmusic - nofeedbacknoise | Negative Ranks | 4 | 3.75 | 15.00 |
| | Positive Ranks | 2 | 3.00 | 6.00 |
| | Ties | 0 | | |
| | Total | 6 | | |
| nofeedanimal - nofeednoise | Negative Ranks | 2 | 3.50 | 7.00 |
| | Positive Ranks | 3 | 2.67 | 8.00 |
| | Ties | 0 | | |
| | Total | 5 | | |
| nofeedbackanimal - nofeedbackmusic | Negative Ranks | 5 | 4.80 | 24.00 |
| | Positive Ranks | 8 | 8.38 | 67.00 |
| | Ties | 0 | | |
| | Total | 13 | | |

A series of Wilcoxon signed ranks test was performed to investigate the performance of participants across the four sound classes without feedback. A two-tailed significance test showed that the classification time threshold for speech (*Mdn* = 29 86), was significantly lower than that for noise (*Mdn* = 83.87), *Z* = 1.99, *p* = .046, music (*Mdn* = 62.61), *Z* = 3.35, *p* =.001, and animal (*Mdn* = 79.58), *Z* = 2.98, *p* = .003. This indicated that the duration of sound clips needed for the participants to classify speech sounds were significantly shorter compared to that for noise, music, and animal sounds. There was no significant difference between music (*Mdn* = 62.61) and noise (*Mdn* = 83.87), *Z* = .943, *p* =.345, noise (Mdn = 83.87) and animal (*Mdn* = 79.58), *Z* = .135, *p* = .863, and animal (*Mdn* = 79.58) and music (*Mdn* = 62.61), *Z* = 1.50, *p* = 1.33. This indicated that there was no significant difference between the duration of sound clips required to classify music and noise, noise and animal, and animal and music as illustrated in Figure 3.1.

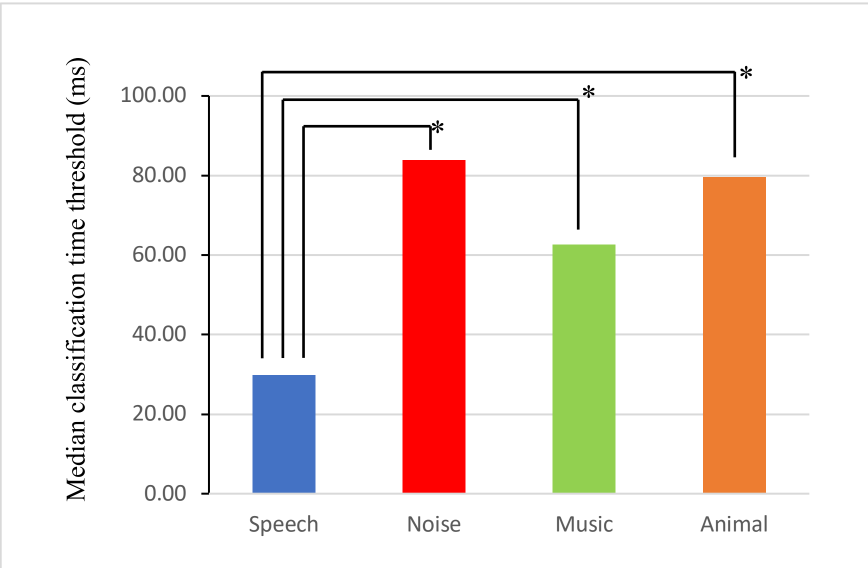



*Figure 3.1.* Classification time thresholds in runs without feedback for all the participants. Speech sounds required the shortest duration to classify in comparison to noise, music, and animal sounds. * $p < 0.05$.

### 3.3.2 Exclusion for classification thresholds without feedback.

Three participants were excluded because of the asymmetry and one participant was excluded because of conductive hearing loss. The classification time thresholds for each sound class was averaged across the two runs before analysis, for each participant. The number of classification time threshold values excluded due to proportion of correct response reaching more than 80% in the last 20 trials are shown in the Table 4. As there were two runs tested for each participant, if a value from one run was excluded, the value from another run was used instead of a mean. One mean value for speech, eight mean values for noise, and two mean values for animal sounds were truly excluded because the values from either of the two runs could not be used. For all the remaining participants, values from at least one run could be used. The individual values that had to be excluded for each sound class across both runs are seen in Table 4. The main problem in testing classification time thresholds without feedback was that proportion of values that had to be excluded for noise (48%) was very high compared to all other sound types. Almost half of all values for noise had to be excluded across Run A and Run B as seen on Table 4.

**Table 4.**

*Number of individual values and proportions excluded from the total of 35 runs for classification time thresholds without feedback across both runs.*

| Sound Class | Speech | Noise | Music | Animal |
|---|---|---|---|---|
| Numbers of values excluded | 2 | 17 | 0 | 1 |
| Proportion of values excluded | 7.4% | 48.6% | 0% | 3.7% |

## 3.4. Part C: Classification time thresholds for with feedback, speech in noise and audiometric assessments.

In this section, the classification time thresholds that were measured with feedback, are analysed. The data from this test are also analysed with speech in noise and audiometric assessments to answer the questions posed in Section 1.10.

### 3.4.1 Distribution of data for mean classification thresholds with feedback.

The classification time thresholds for each sound class was averaged across the two runs before the analysis. As done in Part B, the classification time threshold values which reached more than 80% of correct response reaching in the last 20 trials were excluded. The distribution of the data was checked, and the result is presented in the Table 5. As there were two runs for each participant, if a value from one run was excluded, the value from another run was used instead of a mean. Normal distribution was only found in SRT and the hearing ability occurred with equal probabilities. All the other set of data were either not normally distributed or did not occur with equal probabilities. Therefore, non-parametric statistical tests were used for rest of the analysis.

**Table 5.**

*Series of tests determining distribution of data. The decision made on the null hypothesis based on the statistical tests and significance level is also tabulated.*

| | Null Hypothesis | Statistical Tests | Significance level | Decision |
|---|---|---|---|---|
| $H_0$I | The categories defined by hearing ability, Normal Hearing (NH) and Hearing Impaired (HI) will occur with probabilities of 0.5 and 0.5. | One-Sample Binomial Test | 0.617[1] | Retain null hypothesis |
| $H_0$II | The categories of Sex occur with equal probabilities. | One-Sample Chi-Square Test | 0.024[1] | Reject null hypothesis |
| $H_0$III | The distribution of Age is normal with mean 54.82 and standard deviation 20.870. | One-Sample Kolmogorov-Smirnov Test | 0.005[1] | Reject null hypothesis |
| $H_0$IV | The distribution of Speech reception threshold (SRT) is normal with mean 54.82 and standard deviation 20.870. | One-Sample Kolmogorov-Smirnov Test | 0.200[1, 2] | Retain null hypothesis |
| $H_0$V | The distribution of pure tone average thresholds is normal with mean 24.87 and standard deviation 21.423. | One-Sample Kolmogorov-Smirnov Test | 0.005[1] | Reject null hypothesis |
| $H_0$VI | The distribution of high frequency pure tone average thresholds is normal with mean 36.59 and standard deviation 30.972. | One-Sample Kolmogorov-Smirnov Test | 1.542E-4[1] | Reject null hypothesis |

**Table 6. (continued)**

| | | | | |
|---|---|---|---|---|
| $H_0$VII | The distribution of MeanSpeech is normal with mean 37.35 and standard deviation 19.314. | One-Sample Kolmogorov-Smirnov Test | 1.419E-4[1] | Reject null hypothesis |
| $H_0$VIII | The distribution of MeanNoise is normal with mean 112.36 and standard deviation 141.181. | One-Sample Kolmogorov-Smirnov Test | 1.179E-5[1] | Reject null hypothesis |
| $H_0$IX | The distribution of MeanMusic is normal with mean 92.82 and standard deviation 77.437. | One-Sample Kolmogorov-Smirnov Test | 2.201E-4[1] | Reject null hypothesis |
| $H_0$X | The distribution of MeanAnimal is normal with mean 99.13 and standard deviation 99.719. | One-Sample Kolmogorov-Smirnov Test | 2.181E-8[1] | Reject null hypothesis |

Exact asymptotic significances are displayed. The significance level is .05.
[1]Lillefors Corrected.
[2]This is a lower bound of true significance.

Normal distribution was only found in SRT and the hearing ability occurred with equal probabilities. All the other set of data were either not normally distributed or did not occur with equal probabilities. Therefore, non-parametric statistical tests were used for rest of the analysis.

### 3.4.2 The classification time thresholds for all the participants

Wilcoxon Signed Ranks test was performed to investigate the performance of participants with hearing loss and normal hearing across the four classes of sound. The ranks are shown in Table 6.

**Table 7.**

*Wilcoxon Signed Ranks Test for all the participants in runs with feedback.*

| | Ranks | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Meannoise - | Negative Ranks | 1 | 1.00 | 1.00 |
| Meanspeech | Positive Ranks | 26 | 14.50 | 377.00 |
| | Ties | 0 | | |
| | Total | 27 | | |
| MeanMusic - | Negative Ranks | 1 | 2.00 | 2.00 |
| Meanspeech | Positive Ranks | 33 | 17.97 | 593.00 |
| | Ties | 0 | | |
| | Total | 34 | | |
| MeanAnimal - | Negative Ranks | 5 | 5.40 | 27.00 |
| Meanspeech | Positive Ranks | 29 | 19.59 | 568.00 |
| | Ties | 0 | | |
| | Total | 34 | | |
| MeanMusic - | Negative Ranks | 13 | 18.15 | 236.00 |

**Table 7. (Continued)**

| | | | | |
|---|---|---|---|---|
| Meannoise | Positive Ranks | 16 | 12.44 | 199.00 |
| | Ties | 0 | | |
| | Total | 29 | | |
| MeanAnimal - Meannoise | Negative Ranks | 13 | 16.85 | 219.00 |
| | Positive Ranks | 16 | 13.50 | 216.00 |
| | Ties | 0 | | |
| | Total | 29 | | |
| MeanAnimal - MeanMusic | Negative Ranks | 18 | 17.44 | 314.00 |
| | Positive Ranks | 18 | 19.56 | 352.00 |
| | Ties | 0 | | |
| | Total | 36 | | |

Two-tailed significance test showed that the classification time threshold for speech (*Mdn* = 30.53), was significantly lower than that for noise (*Mdn* = 68.32), $Z = 4.52$, $p < .001$, music (*Mdn* = 63.56), $Z = 5.02$, $p < .001$, and animal (*Mdn* = 65.29), Z = 5.02, $p < .001$. This indicated that the duration of sound clips required for the participants to classify speech sounds was significantly shorter compared to that for noise, music, and animal sounds as seen in Figure 3.2. There was no significant difference between music (*Mdn* = 63.56) and noise (*Mdn* = 68.32), Z = .400, p > .05, noise (*Mdn* = 65.32) and animal (*Mdn* = 65.29), $Z = 0.32$, $p > .05$, and animal (*Mdn* = 65.29), and music (*Mdn* = 63.56) $Z = .299$, $p > .05$. The duration of sound clips required to classify music and noise, animal and noise, and animal and music sounds were similar.

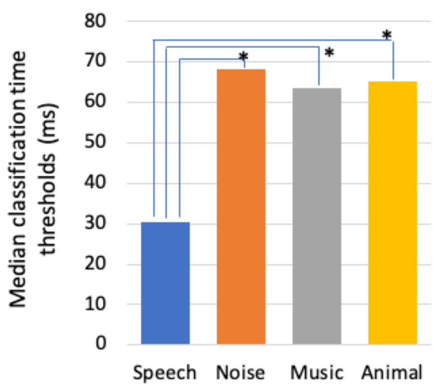



*Figure 3.2.* Classification time thresholds in runs with feedback for different types of sounds for all participants. * p < .001.

### 3.4.3 Classification time thresholds for identifying sounds across the two groups: hearing impaired and normal hearing.

For this analysis, the data were compared by separating the participants based on hearing ability. The two groups were normal hearing and hearing impaired. A Mann-Whitney U test showed that speech classification time thresholds was greater for hearing loss (*Mdn* = 39.75), than for normal hearing (*Mdn* = 26.18), *U*= 45.00, *p* = .001. The result of the ranks is tabulated in Table 7.

**Table 8.**

*Summary of Ranks for participants with normal hearing and hearing impairment.*

| | Hearing ability | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| MeanSpeech | Normal Hearing | 15 | 11.00 | 165.00 |
| | Hearing impaired | 19 | 22.63 | 430.00 |
| | Total | 34 | | |
| MeaNoise | Normal Hearing | 15 | 9.60 | 144.00 |
| | Hearing impaired | 14 | 20.79 | 291.00 |
| | Total | 29 | | |
| MeanMusic | Normal Hearing | 16 | 10.13 | 162.00 |
| | Hearing impaired | 20 | 25.20 | 504.00 |
| | Total | 36 | | |
| MeanAnimal | Normal Hearing | 16 | 16.06 | 257.00 |
| | Hearing impaired | 20 | 20.45 | 409.00 |
| | Total | 36 | | |

The noise classification time threshold was greater for hearing loss (*Mdn* = 128.92), than for normal hearing (*Mdn* = 48.17), *U*= 24.00, *p* < .001. The music classification time threshold was greater for hearing loss (*Mdn* = 99.51), than for normal hearing (*Mdn* = 44.57), *U*= 26.00, *p* < .001. This indicated that the duration of the sound clips needed by participants with hearing loss to classify speech, noise, and music sounds was longer compared to the durations needed by normal hearing participants. For animal sounds, classification time threshold was not significantly different for hearing impaired (*Mdn* =76.81), compared to normal hearing (*Mdn* = 54.55), *U*= 121.00, *p* = .223. This indicated that the durations of the sound clips required to recognise the animal sounds for normal hearing participants was not significantly different compared to participants with hearing loss participants as seen in Figure 3.3.

As reported above, there was significant difference between the classification time thresholds for speech, noise, and music sounds between participants with hearing loss and participants with normal hearing. There was no significant difference in classification of animal sounds. Therefore, the null hypothesis ($H_02$), that there will be no significant differences between the classification time thresholds for participants with hearing impairment and participants with normal hearing for speech, music, animal, and noise sounds, was only supported for animal but not for the others.

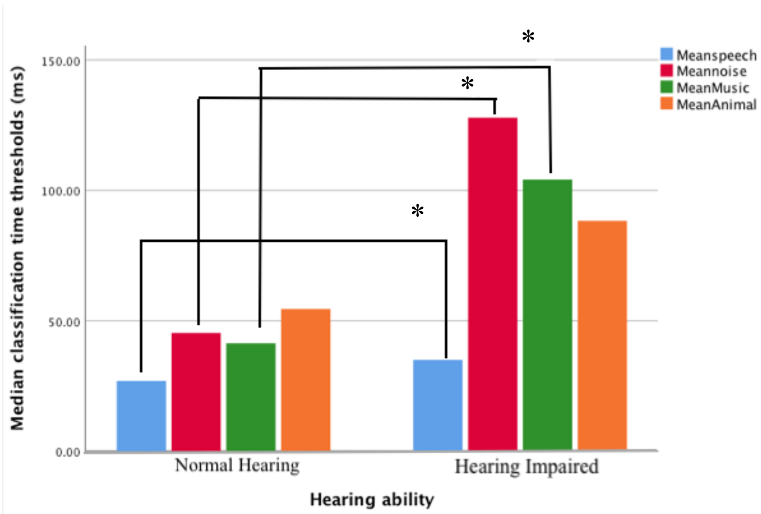



*Figure 3.3.* Classification time thresholds for participants with SNHL and NH. * p < .001.

### 3.4.4 The classification time thresholds within two groups: hearing impaired and normal hearing participants.

For this section, two series of Wilcoxon Signed Ranks tests were performed: one for the normal hearing participants and the other for the hearing-impaired participants. The Wilcoxon Signed Ranks test for normal hearing participants is shown in Table 8. Wilcoxon Signed Ranks test was performed to investigate the performance of participants with normal hearing, across the four classes of sound. Two-tailed significance test showed that the classification time threshold for speech (*Mdn* = 26 .18), was significantly lower than that for noise (*Mdn* = 48.17), $Z = 3.35$, $p < .001$, music (*Mdn* = 44.57), $Z = 3.35$, $p < .001$, and animal (*Mdn* = 54.55), $Z = 3.35$, $p < .001$. This indicated that the duration of sound clips needed for normal hearing participants to classify speech sounds were significantly shorter compared to that for noise, music, and animal sounds. The classification time threshold for animal (*Mdn* = 54.55), was significantly higher than for music (*Mdn* = 44.55) $Z = .299$, $p < .05$, indicating that the duration of sound clips required to classify music sounds were longer compared to that required for animal sounds. There was no significant difference between music (*Mdn* = 44.57) and noise (*Mdn* = 48.17), $Z = .398$, $p = .691$, noise (*Mdn* = 48.17) and animal (*Mdn* = 54.55), $Z = 1.76$, $p = .078$. This indicated that there was no significant difference between the duration of sound clips required to classify music and noise, as well as noise and animal sounds for normal hearing participants as illustrated in Figure 3.4.

**Table 9.**

*Wilcoxon Signed Ranks Test for participants with normal hearing.*

| | Ranks | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| **Meannoise -** | Negative Ranks | 0 | .00 | .00 |
| **Meanspeech** | Positive Ranks | 14 | 7.50 | 105.00 |
| | Ties | 0 | | |
| | Total | 14 | | |
| **MeanMusic - Meanspeech** | Negative Ranks | 1 | 1.00 | 1.00 |
| | Positive Ranks | 14 | 8.50 | 119.00 |
| | Ties | 0 | | |
| | Total | 15 | | |
| **MeanAnimal - Meanspeech** | Negative Ranks | 1 | 1.00 | 1.00 |
| | Positive Ranks | 14 | 8.50 | 119.00 |
| | Ties | 0 | | |
| | Total | 15 | | |
| **MeanMusic - Meannoise** | Negative Ranks | 6 | 11.17 | 67.00 |
| | Positive Ranks | 9 | 5.89 | 53.00 |
| | Ties | 0 | | |
| | Total | 15 | | |
| **MeanAnimal - Meannoise** | Negative Ranks | 4 | 7.25 | 29.00 |
| | Positive Ranks | 11 | 8.27 | 91.00 |
| | Ties | 0[o] | | |
| | Total | 15 | | |
| **MeanAnimal - MeanMusic** | Negative Ranks | 5 | 6.00 | 30.00 |
| | Positive Ranks | 11 | 9.64 | 106.00 |
| | Ties | 0 | | |
| | Total | 16 | | |

The classification time threshold for animal (*Mdn* = 54.55), was significantly higher than for music (*Mdn* = 44.55) $Z = .299$, $p < .05$, indicating that the duration of sound clips required to classify music sounds were longer compared to that required for animal sounds.

There was no significant difference between music (*Mdn* = 44.57) and noise (*Mdn* = 48.17), $Z$ = .398, $p$ = .691, noise (*Mdn* = 48.17) and animal (*Mdn* = 54.55), $Z$ = 1.76, $p$ = .078. This indicated that there was no significant difference between the duration of sound clips required to classify music and noise, as well as noise and animal sounds for normal hearing participants as illustrated in Figure 3.4.

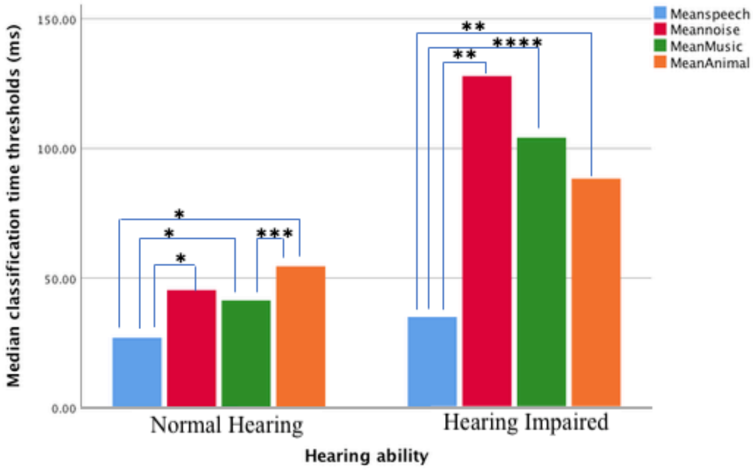



*Figure 3.4.* Classification time thresholds for SNHL and NH participants. The results of two separate series of Wilcoxon Signed Ranks are illustrated. * $p$ = .001, ** $p$ = .002, *** $p$ < .05, **** $p$ < .001.

Another series of Wilcoxon Signed Ranks test was performed to investigate the performance of hearing-impaired participants across the four classes of sound. The results are tabulated in Table 9. Like in participants with normal hearing, the duration of sound clips required to classify speech was shorter compared to all other sound types as seen in Figure 3.4. Two-tailed significance test showed that the classification time threshold for speech (*Mdn* = 39.77), was significantly lower than that for noise (*Mdn* = 128.92), $Z$ = 3.11, $p$ = .002, music (*Mdn* = 99.50), $Z$ = 3.41, $p$ = .001, and animal (*Mdn* = 76.81), $Z$ = 3.14, $p$ = .002. This indicated that the duration of sound clips needed for hearing-impaired participants to classify speech sounds was significantly shorter compared to that for noise, music, and animal sounds. There was no significant difference between music (*Mdn* = 99.50) and noise

(*Mdn* = 68.32), $Z = .282$, $p < .05$, noise (*Mdn* = 128.92) and animal (*Mdn* = 76.81), $Z = 1.10$, $p > .05$, and animal (*Mdn* = 76.81), and music (*Mdn* = 99.50), $Z = 1.12$, $p > .05$. This meant that duration of sound clips required to classify noise and music, and noise and animal sounds were similar. The participants with normal hearing needed the animal sounds to be longer compared to music sounds. However, for participants with hearing impairment, the durations required for classification of music and animal sounds were not significantly different.

**Table 10.**

*Wilcoxon Signed Ranks Test for participants with hearing loss.*

| | Ranks | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| MeanNoiseHL - MeanSpeechHL | Negative Ranks | 1 | 1.00 | 1.00 |
| | Positive Ranks | 12 | 7.50 | 90.00 |
| | Ties | 0 | | |
| | Total | 13 | | |
| MeanMusicHL - MeanSpeechHL | Negative Ranks | 0 | .00 | .00 |
| | Positive Ranks | 19 | 10.00 | 190.00 |
| | Ties | 0 | | |
| | Total | 19 | | |
| MeanAnimalHL - MeanSpeechHL | Negative Ranks | 4 | 4.25 | 17.00 |
| | Positive Ranks | 15 | 11.53 | 173.00 |
| | Ties | 0 | | |
| | Total | 19 | | |
| MeanMusicHL - MeanNoiseHL | Negative Ranks | 7 | 8.14 | 57.00 |
| | Positive Ranks | 7 | 6.86 | 48.00 |
| | Ties | 0 | | |
| | Total | 14 | | |
| MeanAnimalHL - MeanMusicHL | Negative Ranks | 13 | 10.38 | 135.00 |
| | Positive Ranks | 7 | 10.71 | 75.00 |
| | Ties | 0 | | |
| | Total | 20 | | |
| MeanAnimalHL - MeanNoiseHL | Negative Ranks | 9 | 7.78 | 70.00 |
| | Positive Ranks | 5 | 7.00 | 35.00 |
| | Ties | 0 | | |
| | Total | 14 | | |

### 3.4.5 The correlations between the classification time thresholds and pure-tone thresholds.

Several Spearman rank order correlations were carried out to investigate the relationships between PTA thresholds and the four classification time thresholds. Two-tailed tests indicated that there was significant relationship between the PTA thresholds, and speech classification $r_S(34) = .587$, $p < .001$, between the PTA and noise classification thresholds

$r_S(29) = .667$, $p < .001$, and between the PTA thresholds and music classification thresholds $r_S(36) = .735$, $p < .001$. This indicated that the higher the PTA threshold of a participants, the longer duration of sound it took for the participants to classify speech, noise, and music sounds. However, there was no significant correlation between animal classification thresholds and PTA thresholds $r_S(36) = .278$, $p = .101$, indicating that the PTA thresholds did not relate to their performance in classifying animal sounds. Refer to Table 10 for the correlations.

**Table 11.**

*The Spearman rank order correlations between PTA thresholds and classifications time thresholds.*

| | | PTA | Mean speech | Mean noise | Mean music | Mean animal |
|---|---|---|---|---|---|---|
| **PTA** | Correlation Coefficient | 1.000 | .587** | .667** | .735** | .278 |
| | Sig. (2-tailed) | . | .000 | .000 | .000 | .101 |
| | N | 36 | 34 | 29 | 36 | 36 |
| **Mean speech** | Correlation Coefficient | .587** | 1.000 | .562** | .674** | .331 |
| | Sig. (2-tailed) | .000 | . | .002 | .000 | .056 |
| | N | 34 | 34 | 27 | 34 | 34 |
| **Mean noise** | Correlation Coefficient | .667** | .562** | 1.000 | .551** | .358 |
| | Sig. (2-tailed) | .000 | .002 | . | .002 | .056 |
| | N | 29 | 27 | 29 | 29 | 29 |
| **Mean music** | Correlation Coefficient | .735** | .674** | .551** | 1.000 | .303 |
| | Sig. (2-tailed) | .000 | .000 | .002 | . | .072 |
| | N | 36 | 34 | 29 | 36 | 36 |
| **Mean animal** | Correlation Coefficient | .278 | .331 | .358 | .303 | 1.000 |
| | Sig. (2-tailed) | .101 | .056 | .056 | .072 | . |
| | N | 36 | 34 | 29 | 36 | 36 |

**Correlation is significant at the 0.01 level (2-tailed).

In summary, there was significant correlation between PTA thresholds and the classification time thresholds for speech, noise, and music sounds, and but there was no correlation between in classification thresholds for animal sounds and PTA thresholds. Therefore, the null hypothesis ($H_02$), that the classification time thresholds will not be correlated to pure tone thresholds was not supported for speech, music, animal, and noise sounds, and was only supported for animal sounds.

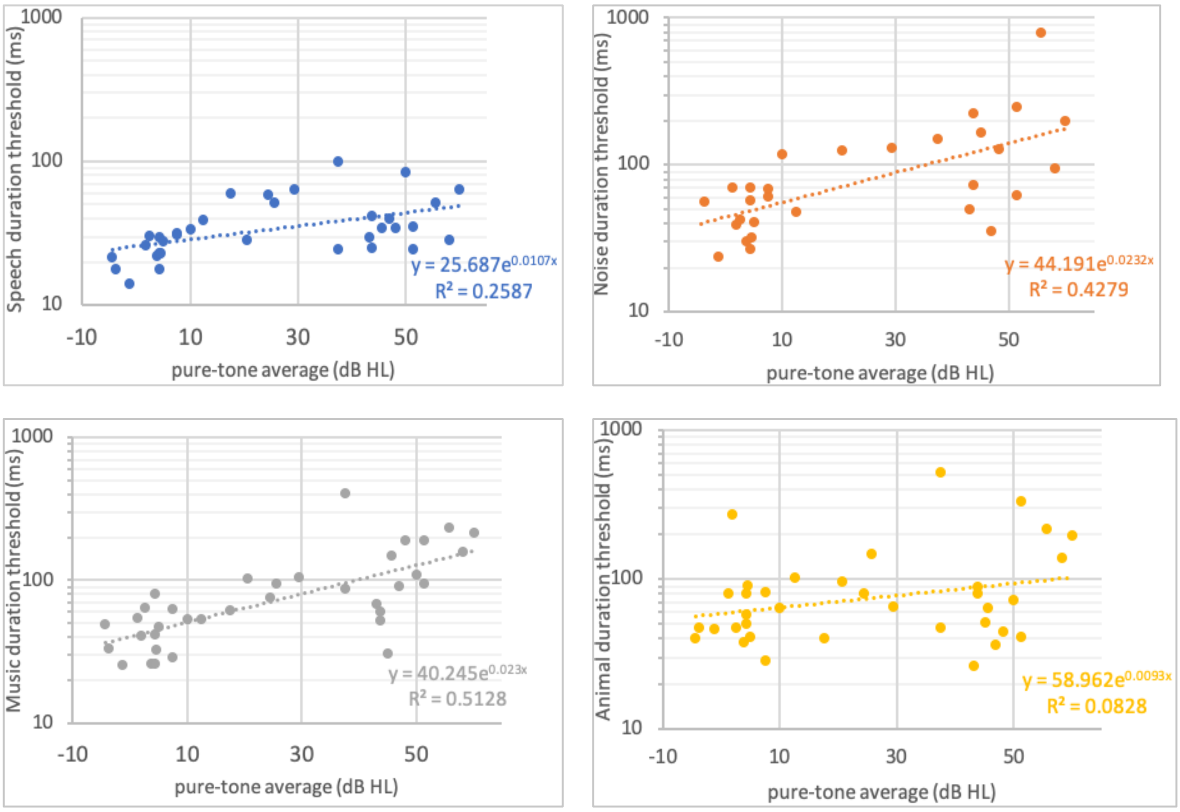



*Figure 3.5.* Correlations between the PTA thresholds and classification time thresholds. The y-axis is a logarithmic and the line of best fit is exponential.

**3.4.6 The correlation between SRT in noise and PTA thresholds.**

Spearman rank-order correlation was carried out to investigate the relationships between PTA thresholds and SRT. Two-tailed test showed that there was significant correlation between PTA and SRT $r_S(36) = .791$, $p <. 001$ as shown in Table 11. This indicated that the worse/higher the PTA, the better the single-to-noise ratio a participant needs for recognising speech in noise. A linear line of best fit was plotted and 63.02% of variance in SRT was accounted for by the variance in PTA thresholds ($R^2 = 0.6302$) as seen in Figure 3.6. Therefore, the null hypothesis ($H_0 3$) that there will be no correlation between SRT, and PTA thresholds was not supported.

**Table 12.**

*Correlation between the PTA thresholds and SRT.*

| | | PTA | SRT |
|---|---|---|---|
| PTA | Correlation Coefficient | 1.000 | .791** |
| | Sig. (2-tailed) | . | .000 |
| | N | 36 | 36 |
| SRT | Correlation Coefficient | .791** | 1.000 |
| | Sig. (2-tailed) | .000 | . |
| | N | 36 | 36 |

** Correlation is significant at the 0.01 level (2-tailed)

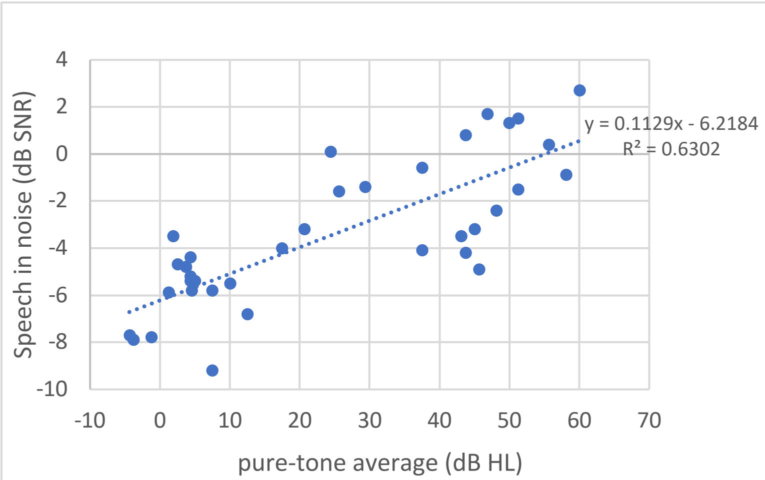



*Figure 3.6.* Correlation between PTA thresholds and SRT. An equation defining the linear best line of fit and the $R^2$ value are also shown.

**3.4.7 The correlations between the classification time thresholds and SRTs**

A series of Spearman rank order correlations were carried out to investigate the relationships between SRT and the four classification time thresholds. Two-tailed test indicated a significant relationship between speech classification thresholds and SRTs $r_S(34) = .535$, $p = .001$, between SRT and noise classification thresholds $r_S(29) = .521$, $p = .004$, and between music classification thresholds and SRT $r_S(36) = .680$, $p < .001$. This indicated that the higher the speech reception in noise, the longer the duration of sound clips it took for the participant to classify noise, music, and animal sounds. However, there was no significant correlation between animal classification thresholds and SRT $r_S(36) = .308$, $p = .067$, indicating that the SRT did not relate to their performance in classifying animal sounds. Refer to Table 12 for the correlations. The best line of fit was exponential in shape. The equitation for best line of fit and $R^2$ values for each correlation are shown in Figure 3.7. The y-axis is plotted in logarithmic scale which makes the best line of fit appear linear.

**Table 13**

*Spearman rank order correlation between SRTs and classifications time thresholds.*

| | | SRT | Mean speech | Mean noise | Mean music | Mean animal |
|---|---|---|---|---|---|---|
| **SRT** | Correlation Coefficient | 1.000 | .535** | .521** | .680** | .308 |
| | Sig. (2-tailed) | . | .001 | .004 | .000 | .067 |
| | N | 36 | 34 | 29 | 36 | 36 |
| **Meanspeech** | Correlation Coefficient | .535** | 1.000 | .562** | .674** | .331 |
| | Sig. (2-tailed) | .001 | . | .002 | .000 | .056 |
| | N | 34 | 34 | 27 | 34 | 34 |
| **Meannoise** | Correlation Coefficient | .521** | .562** | 1.000 | .551** | .358 |
| | Sig. (2-tailed) | .004 | .002 | . | .002 | .056 |
| | N | 29 | 27 | 29 | 29 | 29 |
| **Mean music** | Correlation Coefficient | .680** | .674** | .551** | 1.000 | .303 |
| | Sig. (2-tailed) | .000 | .000 | .002 | . | .072 |
| | N | 36 | 34 | 29 | 36 | 36 |
| **Mean animal** | Correlation Coefficient | .308 | .331 | .358 | .303 | 1.000 |
| | Sig. (2-tailed) | .067 | .056 | .056 | .072 | . |
| | N | 36 | 34 | 29 | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

In summary, there was significant correlation between SRT thresholds and the classification time thresholds for speech, noise, and music sounds. There was no significant correlation between in classification thresholds for animal sounds and SRT thresholds. Therefore, the null hypothesis ($H_0$4), that the classification time thresholds will not be correlated to SRT thresholds was not supported for speech, music, animal, and noise sounds, was only supported for animal sounds.

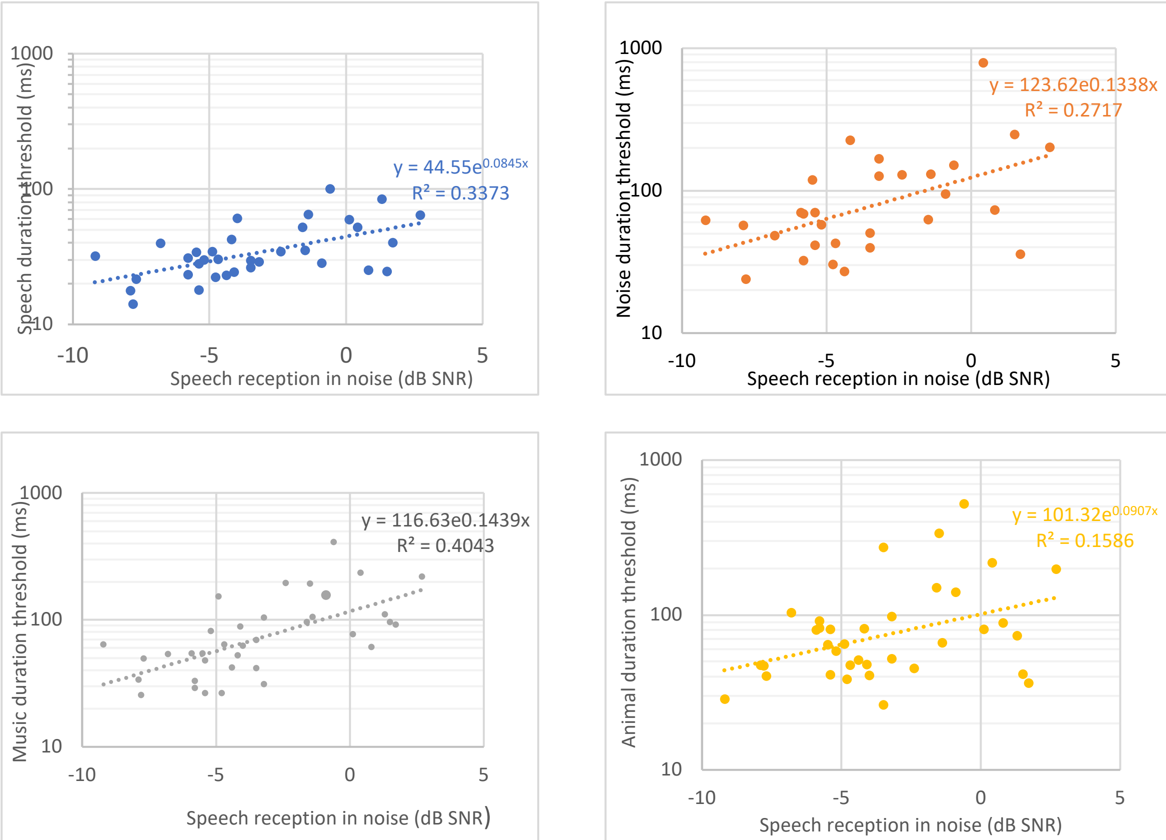



*Figure 3.7.* Correlations between the SRT and classification time thresholds. The y-axis is logarithmic, and the lines of best fit are exponential as defined by their respective equations.

### 3.4.8 The correlation between age and PTA thresholds

Spearman rank-order correlation was carried out to investigate the relationships between PTA thresholds and age. Two-tailed test showed that there was significant correlation between PTA and age $r_S(36) = .669$, $p < .001$ as shown in Table 13.

**Table 14.**

*The correlation between the pure-tone average thresholds and age.*

| | | PTA | Age |
|---|---|---|---|
| PTA | Correlation Coefficient | 1.000 | .669** |
| | Sig. (2-tailed) | . | .000 |
| | N | 36 | 36 |
| Age | Correlation Coefficient | .669** | 1.000 |
| | Sig. (2-tailed) | .000 | . |
| | N | 36 | 36 |

** Correlation is significant at the 0.01 level (2-tailed)

This indicated that the higher the age of a participant the worse/higher the PTA threshold of a participant. A linear best fit of line was formed and 50% of variance was accounted for ($R^2 = 0.499$) as seen in Figure 3.8. Therefore, the null hypothesis ($H_05$) that there will be no correlation between age and PTA thresholds was not supported.

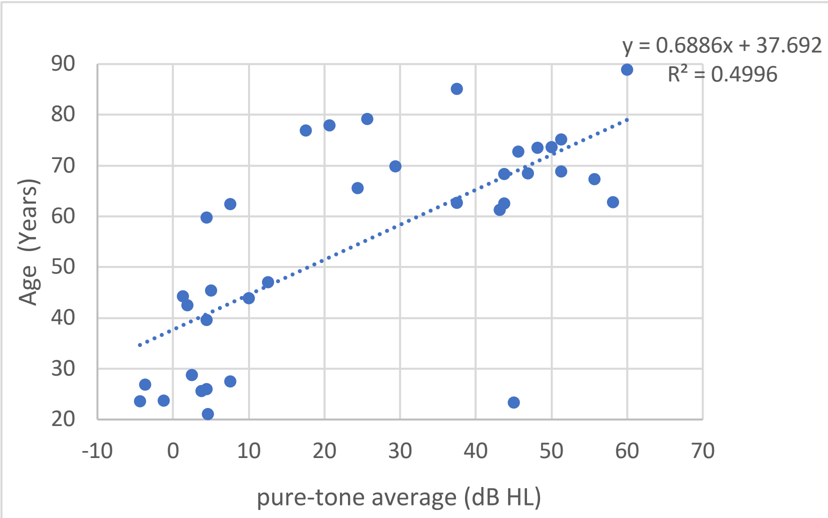



*Figure 3.8.* Correlation between the pure-tone thresholds and age. An equation defining the linear line of fit and the $R^2$ value are also shown.

**3.4.9 The correlations between age and Classification time thresholds**

A series of Spearman rank-order correlations were carried out to investigate the relationships between age and the four classification time thresholds. Two-tailed test indicated significant relationship between age, and speech classification thresholds $r_S(34) = .738$, $p < .001$, between age and noise classification thresholds $r_S(29) = .606$, p < .001, between age and music classification thresholds $r_S(36) = .822$, $p < .001$, and between age and animal classification thresholds $r_S(36) = .379$, $p < .005$. This indicated that the higher the age of a participant, the longer the duration of sound clips it took for the participant to classify speech, noise, music, and animal sounds. Refer to Table 14 for the correlations. The line of fit was added and the equation and $R^2$ value for each correlation is plotted in Figure 3.9. Although, the line of fit is exponential, it appears linear because the y-axis is logarithmic.

**Table 15.**

*Spearman rank order correlations between age and classifications time thresholds.*

| | | Age | Mean speech | Mean noise | Mean music | Mean animal |
|---|---|---|---|---|---|---|
| **AGE** | Correlation Coefficient | 1.000 | .738** | .606** | .822** | .379* |
| | Sig. (2-tailed) | . | .000 | .000 | .000 | .023 |
| | N | 36 | 34 | 29 | 36 | 36 |
| **Meanspeech** | Correlation Coefficient | .738** | 1.000 | .562** | .674** | .331 |
| | Sig. (2-tailed) | .000 | . | .002 | .000 | .056 |
| | N | 34 | 34 | 27 | 34 | 34 |
| **Meannoise** | Correlation Coefficient | .606** | .562** | 1.000 | .551** | .358 |
| | Sig. (2-tailed) | .000 | .002 | . | .002 | .056 |
| | N | 29 | 27 | 29 | 29 | 29 |
| **Mean music** | Correlation Coefficient | .822** | .674** | .551** | 1.000 | .303 |
| | Sig. (2-tailed) | .000 | .000 | .002 | . | .072 |
| | N | 36 | 34 | 29 | 36 | 36 |
| **Mean animal** | Correlation Coefficient | .379* | .331 | .358 | .303 | 1.000 |
| | Sig. (2-tailed) | .023 | .056 | .056 | .072 | . |
| | N | 36 | 34 | 29 | 36 | 36 |

**. Correlation is significant at the 0.01 level (2-tailed).

In summary, there were significant correlations between the age and classification time thresholds for speech, noise, and music, animal sounds. Therefore, the null hypothesis ($H_0$6), that there will be no correlation between age and classification time thresholds was supported.

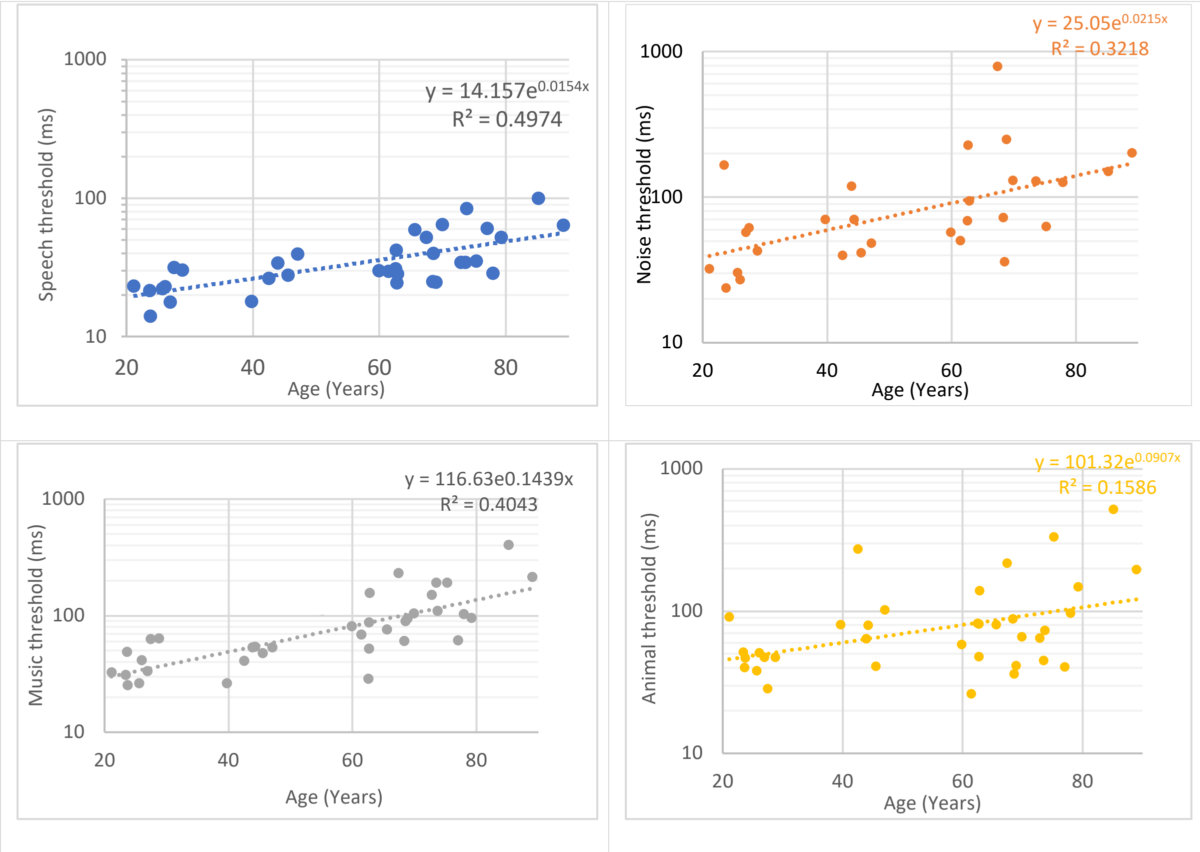



*Figure 3.9.* Correlations between the PTA and age. The y-axis is logarithmic, and the line of fit is exponential as defined by their respective equations.

### 3.4.10 Practise effect: Run A and Run B for classification time thresholds with feedback

The data for both runs were individually analysed to investigate the distribution. It was found that the of distribution of the data was not normal for all variables. See Table 15 for the tests carried. Therefore, a non-parametric test, Wilcoxon Signed Ranks test, was performed as seen in Table 16. Two-tailed significance test showed that the speech classification time threshold for Run A (*Mdn* = 33.94), was not significantly different compared to the Run B (*Mdn* = 26.19), $Z = .956$, $p = .339$. The noise classification time threshold for Run A (*Mdn* = 64.83), was not significantly different compared to the Run B (*Mdn* = 67.29), $Z = .644$, p = .520. The animal classification time threshold for Run A (*Mdn* = 76.23), was not significantly different compared to the Run B (*Mdn* = 48.25), $Z = .767$, $p = .443$.

**Table 16**

*One-Sample Kolmogorov-Smirnov Test for Run A and Run B of time classification thresholds with feedback.*

| | Null Hypothesis | Significance level | Decision |
|---|---|---|---|
| $H_0I$ | The distribution of SpeechA is normal with mean 37.90 and standard deviation 19.364. | .007[1] | Reject null hypothesis |
| $H_0II$ | The distribution of NoiseA is normal with mean 81.20 and standard deviation 69.556. | .001[1] | Reject null hypothesis |
| $H_0III$ | The distribution of MusicA is normal with mean 98.22 and standard deviation 80.493. | .001[1] | Reject null hypothesis |
| $H_0IV$ | The distribution of AnimalA is normal with mean 103.47 and standard deviation 106.418. | .000[1] | Reject null hypothesis |
| $H_0V$ | The distribution of SpeechB is normal with mean 34.87 and standard deviation 19.718. | .000[1] | Reject null hypothesis |
| $H_0VI$ | The distribution of NoiseB is normal with mean 153.39 and standard deviation 293.197. | .000[1] | Reject null hypothesis |
| $H_0VII$ | The distribution of MusicB is normal with mean 79.18 and standard deviation 88.224. | .001[1] | Reject null hypothesis |
| $H_0VIII$ | The distribution of AnimalB is normal with mean 87.31 and standard deviation 99.719. | .000[1] | Reject null hypothesis |

Asymptotic significances are displayed. The significance level is .05.
[1]Lillefors Corrected.

**Table 17.**

*Wilcoxon signed ranks for Run A and Run B of classification time thresholds with feedback.*

| | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| SpeechB - SpeechA | Negative Ranks | 16 | 17.44 | 279.00 |
| | Positive Ranks | 14 | 13.29 | 186.00 |
| | Ties | 0c | | |
| | Total | 30 | | |
| NoiseB - NoiseA | Negative Ranks | 8d | 9.88 | 79.00 |
| | Positive Ranks | 11 | 10.09 | 111.00 |
| | Ties | 0 | | |
| | Total | 19 | | |
| MusicB - MusicA | Negative Ranks | 21 | 18.95 | 398.00 |
| | Positive Ranks | 11 | 11.82 | 130.00 |
| | Ties | 0 | | |
| | Total | 32 | | |
| AnimalB - AnimalA | Negative Ranks | 18 | 10.61 | 191.00 |
| | Positive Ranks | 7 | 19.14 | 134.00 |
| | Ties | 0 | | |
| | Total | 25 | | |

This meant that the duration of sound clips required to classify speech, noise, and animal sounds for Run A and Run B were similar and there was no practise effect. In contrast, the classification time threshold for music was significantly lower for Run B (*Mdn* = 56.66), compared to Run A (*Mdn* = 78.11), $Z = 2.50$ $p = .012$. The duration of sound clips required to classify music was significantly lower for Run B compared to Run A, which indicated that there could have been practise effect for this particular sound class. Therefore, in runs with feedback, the null hypothesis ($H_0 7$) that there will be no significant difference between Run A and Run B of classification time thresholds was supported for speech, noise, and animal sounds but for not for music sound.

We further investigated if there was significant difference between Run A and Run B of music for classification time thresholds in runs without any feedback. The data for Run A and Run B of classification time thresholds without feedback were analysed to investigate the distribution. It was found that the of distribution of the data was not normal for music sounds in Run B. Therefore, Wilcoxon Signed Ranks test was used. Two-tailed significance test showed that the music classification time threshold for Run A (*Mdn* = 82.13), was not significantly different compared to the Run B (*Mdn* = 54.90), $Z = .923$, $p = .356$. This meant that the duration of sound clips required to classify speech, noise, and animal sounds for Run A and Run B were similar and there was no practise effect. Therefore, in runs without feedback, the null hypothesis ($H_0 7$) that there will be no significant difference between Run A and Run B of classification time thresholds was supported for music sounds.

**3.4.11 The classification time thresholds for runs with feedback and without feedback.**

This part of the study only included the nine participants who participated in both tests with the feedback and without feedback. There were two runs for each condition. The mean for the Run A and Run B was calculated for each sound class before the analysis. Normal distribution was checked using One-Sample Kolmogorov-Smirnov test as seen in Table 17.

The data for speech classification with no feedback was not normal. Therefore, Wilcoxon Signed Ranks Tests, as seen in Table 18, was performed. There was no significant difference between classification time thresholds for speech sounds with feedback (*Mdn* = 26.87), and without feedback (*Mdn* = 20.11), $Z = .980$, $p = .327$. There was no significant difference between classification time thresholds for noise sounds with feedback (*Mdn* = 66.95), and without feedback (*Mdn* = 43.12), $Z = .365$, $p = .715$. There was no significant difference between classification time thresholds for music sounds with feedback (*Mdn* = 62.16), and without feedback (*Mdn* = 69.86), $Z = 1.96$, $p = .051$. There was no significant difference between classification time thresholds for animal sounds with feedback (*Mdn* = 67.87), and without feedback (*Mdn* = 77.86), $Z = 1.13$, $p = .260$.

In summary, for participants who participated in tests with both feedback and without feedback runs, the median duration of sound clips required to classify each of the four classification time thresholds was not significantly different in both tests. Therefore, the null hypothesis ($H_0 8$) that there will be no significant difference between classification time thresholds with feedback without feedback was supported.

**Table 18.**

*One-Sample Kolmogorov-Smirnov Test for classification time thresholds with feedback and without feedback.*

| | Null Hypothesis | Significance level | Decision |
|---|---|---|---|
| $H_0I$ | The distribution of no feedback speech is normal with mean 27.22 and standard deviation 12.685. | .005c [1] | Reject null hypothesis |
| $H_0II$ | The distribution of no feedback noise is normal with mean 72.82 and standard deviation 51.168. | .200[1, 2] | Retain null hypothesis |
| $H_0III$ | The distribution of no feedback music is normal with mean 70.47 and standard deviation 30.953. | .200[1, 2] | Retain null hypothesis |
| $H_0IV$ | The distribution of no feedback animal is normal with mean 75.10 and standard deviation 53.564. | .176[1] | Retain null hypothesis |
| $H_0V$ | The distribution of feedback speech is normal with mean 27.22 and standard deviation 12.685. | . 111[1,] | Retain null hypothesis |
| $H_0VI$ | The distribution of feedback noise is normal with mean 72.82 and standard deviation 51.168. | .200[1, 2] | Retain null hypothesis |
| $H_0VII$ | The distribution of feedback music is normal with mean 70.47 and standard deviation 30.953. | .200[1, 2] | Retain null hypothesis |
| $H_0VIII$ | The distribution of feedback animal is normal with mean 75.10 and standard deviation 53.564. | .062 [1] | Retain null hypothesis |

Asymptotic significances are displayed. The significance level is .05.
[1]Exact significance is displayed for this test.
[1]Lillefors Corrected.
[2]This is a lower bound of true significance.

**Table 19.**

*Wilcoxon Signed Ranks for classification time thresholds with feedback and without feedback.*

| | | N | Mean Rank | Sum of Ranks |
|---|---|---|---|---|
| Feedback Speech - | Negative Ranks | 2 | 5.50 | 11.00 |
| No Feedback Speech | Positive Ranks | 6 | 4.17 | 25.00 |
| | Ties | 0 | | |
| | Total | 8 | | |
| Feedback Noise - | Negative Ranks | 2 | 2.00 | 4.00 |
| No Feedback Noise | Positive Ranks | 2 | 3.00 | 6.00 |
| | Ties | 0 | | |
| | Total | 4 | | |
| Feedback Music - | Negative Ranks | 7 | 5.57 | 39.00 |
| No Feedback Music | Positive Ranks | 2 | 3.00 | 6.00 |
| | Ties | 0 | | |
| | Total | 9 | | |
| Feedback Animal - | Negative Ranks | 6 | 5.33 | 32.00 |
| No Feedback Animal | Positive Ranks | 3 | 4.33 | 13.00 |
| | Ties | 0 | | |
| | Total | 9 | | |

### 3.4.12 Exclusion for classification time thresholds with feedback.

Four participants were excluded because of significant air-bone gap. Three participants were excluded because of significant asymmetry. One participant was excluded because of both air bone gap and significant asymmetry. Only one run of classification time thresholds was carried out for six participants due to time constraints. The number of values excluded from each run due to proportion of correct response reaching more than 80% in the last 20 trials are shown in the Table 19. There were only two mean values for speech and seven mean values for noise which were truly excluded, because the values from either of the two runs could not be used or were unavailable. For all the remaining participants data from at least one run could be used.

**Table 20.**

*Number of individual values and proportions from the total of 68 runs excluded across both runs for classification time thresholds with feedback.*

| Sound Class | Speech | Noise | Music | Animal |
|---|---|---|---|---|
| Numbers removed | 5 | 19 | 0 | 7 |
| percentage removed | 7.4% | 27.9% | 0% | 10.3% |

### 3.4.13 Outliers for Classification time thresholds with feedback.

Using the descriptive statistics in SPSS Statistics (IBM Corp, Armonk, NY, USA) outliers were searched. The interquartile range was multiplied by 1.5 and 3. The resulting values above were considered outliers. The number of outliers for each sound class are listed in Table 20. They were not excluded from the non-parametric analysis.

**Table 21.**

*Number of significant outliers in each classification time thresholds*

| Sound Class | MeanSpeech | MeanNoise | MeanMusic | MeanAnimal |
|---|---|---|---|---|
| IQR x1.5 | 2 | 0 | 2 | 3 |
| IQR x13 | 1 | 1 | 1 | 2 |
| Total | 3 | 1 | 3 | 5 |

**3.5 Summary of results.**

1. Results were not statistically analysed for the pilot study.
2. For classification time thresholds without feedback speech sounds took the shortest time to classify compared to other sounds. However, there were very high number of individual values that needed to be excluded due to unrealistically high number of correct responses and in some cases the absence of reversals.
3. This problem was reduced with addition of feedback in Part C.
4. Classification time thresholds with feedback for speech sounds also took the shortest time to classify compared to all other sounds.
5. Classification time thresholds for participants with hearing loss was significantly higher compared to that for participants with normal hearing for speech, noise and music sounds but not animal sounds.
6. There were significant positive correlations between classification time threshold and pure tone thresholds for speech, noise, and music sounds but not animal sounds.
7. There was significant correlation between PTA thresholds and SRTs in noise.
8. There was significant positive correlation between classification time threshold and speech reception thresholds in noise for speech, noise, and music sounds but not animal sounds.
9. There was significant correlation between age and PTA thresholds.
10. There was significant correlation between age and classification time thresholds for speech, noise, music, and animal sounds.
11. There was no significant difference between classification time thresholds with feedback and without feedback.
12. Music sounds took shorter time to classify in Run A compared to Run B in runs with feedback sounds. This difference was not observed in runs without feedback.

# 4. Discussion

## 4.1 Overview

The purpose of this study was to build on the research by Obert and Tchorz (2018) and improve methods that can reliably measure the minimum time needed to classify the four major categories of sounds. In addition, the study aimed to investigate whether the classification time thresholds would correlate with PTA thresholds and SRTs. In this chapter, the questions posed in the Section 1.10 and the hypotheses related with those questions are discussed. The applications and limitations of the current study and the possible directions future studies are also discussed in this chapter. The rationale for methods used and the changes needed to be made to the methods are detailed in the Method section. They are only discussed here when discussing limitation of the study.

## 4.2. Classification time thresholds for participants with hearing impairment and normal hearing and their correlation with PTA thresholds.

The first null hypothesis ($H_01$) was that there will be no significant differences between the classification time thresholds for participants with hearing impairment and participants with normal hearing. It was supported for animal sounds but not supported for speech, noise, and music sounds. The classification time thresholds were higher for participants with hearing impairment compared to those with normal hearing for speech, noise, and music sounds. This meant that the duration of the sounds required by the participants with SNHL to classify speech, noise, and music sounds were significantly longer compared to that required by participants with normal hearing.

Previous studies have found that participants with SNHL have impaired temporal processing indicated by poorer performance in gap detection and TMTF tests. For example, in gap detection tests, the gap duration in noise needed to be longer for detection by those with SNHL compared to individuals with NH (Florentine & Buss, 1984). Most stimuli used in the current study for measurement of classification time thresholds contained slow, random temporal fluctuations from one movement to other. Participants with SNHL are poorer at processing these fluctuations compared to those with NH. This can cause the participants with SNHL to miss important acoustic cues in the temporal domain (Moore, 2008).

Hearing loss affects frequency selectivity which is important in analysis of the spectral composition of sounds. An accurate analysis of spectral composition of sounds is required for the perception of music and speech sounds. In individuals with SNHL, frequency selectivity is reduced which leads to difficulty in discrimination between different vowel sounds as well as different musical instruments (Moore, 2013). Strelcyk and Dau, (2009) have shown that the perception of temporal fine structure is impaired in participants with SNHL. The impaired ability to follow temporal structure along with other processing difficulties like frequency selectivity and frequency discriminations may be some of the reasons that makes the participants with SNHL slower in classifying short sounds. One of the reasons, why participants with SNHL required longer duration to classify sounds, may be that they needed more acoustic information to confidently make the judgements on sound class. Previous studies that have used psychoacoustical tests like gap detection and TMTF with shorter stimuli like tones and noises have shown that individuals with SNHL have reduced temporal resolution (Glasberg, Moore, & Bacon, 1987; Schneider et al., 1994). The data from the current study showed that the temporal resolution is poorer in individuals with SNHL even for stimuli with relatively longer duration and complex information.

More interestingly, the null hypothesis ($H_02$) that the classification time thresholds will not be correlated to pure tone thresholds was only supported for animal sounds, as the data showed that there was significant positive correlation between pure tone thresholds and time classification thresholds for speech, noise, music but not for animal sounds. This meant that the higher the PTA thresholds (worse hearing), the higher the durations of the sounds required by the participants to classify speech, music, and noise sounds. Even though PTA thresholds are poor predictors of temporal processing abilities like gap detection thresholds, significant hearing impairment is associated with poor temporal processing abilities (Moore, 2013). Likewise, frequency selectivity is poor in people with HI, in part due to broader filters but there is small but significant correlation between PTA thresholds and broadening of the filters. Moreover, higher thresholds tend to be associated with broader filters (Tyler, 1986). The performance in frequency discrimination tasks is also generally worse in individuals with SNHL (Grant, 1987; Zurek & Formby, 1981). The psychoacoustic tests investigating temporal processing, frequency selectivity, and frequency discrimination measure specific auditory processing abilities. Thus, it may be the reason why they show weak correlations or no significant correlation with PTA thresholds. Whereas, classifying short sounds requires the combined processing abilities including perception of temporal fine structure, temporal processing, frequency selectivity and frequency discrimination. Therefore, classification time thresholds measuring combined processing abilities may have provided stronger correlation between hearing abilities and classification time thresholds for speech, noise, and music sounds.

For all participants the speech sounds took significantly shorter duration to classify compared to all other sound types. This pattern was seen in both groups - hearing impaired and normal hearing - and in both runs with feedback and without feedback. In the current study the median classification time threshold for speech was 39.75 ms for participants with SNHL and 26.18 ms for participants with normal hearing. Previous studies have shown individual can detect and identify speech sounds in a very short duration. Gray (1942), showed that vowel sounds could be recognised by some normal hearing individuals at duration of as little as 3 ms. Moradi et al., (2013) showed that it took 16.67 ms for the normal hearing individuals to correctly identify consonants sounds. Isnard, Taffou, Viaud-Delmon, and Suied (2016), showed that even when the quality of sounds was highly impoverished, speech sound was recognised exceptionally well compared to animal sounds or instrumental music. Agus, Suied, Thorpe, and Pressnitzer (2012), reported that speech sounds are recognised faster than musical sounds. The use of different procedure, method, and stimuli for testing classification and recognition of sounds may be the reason for variations in reported time required to classify and recognise sounds. There are no standard procedures, methods, or stimuli used to test classification of sounds yet.

Moore (2003), points out that acoustic patterns of short speech sounds may not be perceived as a series of distinct events but rather as group of acoustic patterns perceived as a specific sound. Furthermore, Moore (2013), suggested that speech may be detected from a small portion of speech spectrum, whereas recognition may require audibility of wide frequency range. In this study the participants, only had to detect the sounds and classify it and did not require recognition of a specific word, vowel, or consonant.

In addition, some researchers argue that speech sounds may be processed differently through a “speech mode”. The supporter of this idea provide evidence for this is presenting cases where speech may require special processing compared to other sounds. Liberman, Cooper, Shankweiler, and Studdert-Kennedy (1967), highlighted that perception of phonemes has a complex relation with acoustic pattern. An acoustic pattern for a phoneme may be influenced by the acoustic pattern of neighbouring phonemes. An acoustic pattern for a vowel sounds never has a corresponding acoustic pattern of its own, as its acoustic pattern is always influenced by the acoustics patterns of consonant sounds before or after it. Therefore, the study argued that a special decoder may be needed for phoneme perception. The existence of special decoder is based on the theory that phonemes are a fundamental unit of speech perception, however, this is disputed by some (Mehler & Hayes, 1981; Plomp, 2002). Another evidence for “speech mode” is categorical perception. Categorical perception is a pheromone where a small progressive change in acoustic signal does not change perception of the sound within a category, but an equally small change across categories changes the perception of sound. This was shown by Liberman et al., (1967), where they progressively changed the synthetic acoustic signal of the syllable /bi/ to /di/. They reported that with the progressive change in acoustic signal there was no slow change in perception, rather the perception suddenly "jumped" from one perceptual category (/bi/) to another (/di/). Categorial perception is mostly seen in speech sounds and are rare in other sound types, further supporting the theory that speech sound is processed differently to others.

There are also studies that show specific regions in the brains are activated when listening to speech sounds. Studies have shown that the upper part of superior temporal sulcus and anterior temporal lobes of the brain are strongly activated when listening to human speech sounds but not when listening to other non-speech sounds (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Bethmann, & Brechmann, 2014). Moreover, studies using EEG (electroencephalography) have shown that peak activity in the speech specific area in the brain occurs around 70ms for speech sounds, whereas 100ms for musical instruments (Murray, Camen, Gonzalez-Andino, Bovet, & Clarke 2006; Rigoulot, Pell, & Armony, 2015). Ogg et al., (2017) argues that because participants can classify sounds in even shorter duration, it could suggest that some relevant acoustic information is extracted very early on. This might help with more accurate processing later as suggested by EEG, that might include characteristics of sound precepts requiring longer temporal windows.

Moreover, the human auditory system can detect sounds between the frequency range of 0.02 and 20 kHz and the most important speech signals lie between frequency range of 0.10 and 5 kHz. It has been shown that humans ears are most sensitive in detection of sounds in the speech frequency range (Fletcher & Munson, 1933). The sound that we are most exposed to in daily life is speech sounds. In 1 second of rapid speech, there can be up to 30 phonemes (Moore, 2013). Therefore, faster processing of speech compared to other sounds should not be unexpected in the auditory system that may be specialised in the detections of speech above all other sounds.

**4.4 Correlation between SRT and PTA thresholds.**

The null hypothesis that the SRTs will not correlate with PTA thresholds is not supported as the data showed that there was a strong correlation between PTA and SRTs. This means that the higher the PTA thresholds (higher the degree of hearing loss), the higher the SRTs (harder it is for participants to understand speech in noise). The finding is consistent with other studies that have also shown that there is a correlation PTA thresholds and speech understanding in noise (Ching, Dillon, & Byrne, 1998; Smoorenburg, 1992; Lutman, 1991). It is widely accepted that people with SNHL struggle to understand speech in noisy environment (Hopkins & Moore, 2011).

However, there are also study that have reported that similar performance in understanding speech in noise may be found in people with very different pure-tone thresholds and vice-versa (Duquesnoy, 1983; Middelweerd, Festen, & Plomp 1990). A lot of variability in SRT values were not accounted for by the variability in the audiometric thresholds and the studies suggest that the significant variance in SRT thresholds can be accounted for by performance in tests of frequency selectivity or temporal resolution (Horst, Javel, & Farley, 1990; Tyler et al., 1982). The study by Glasberg and Moore (1989) showed that the performance in tests of frequency discrimination of complex tones and pure tone, as well as performance in temporal resolution test involving detection of temporal gaps in noise correlated more highly to SRT than the PTA thresholds. Therefore, it suggested that the ability to understand speech in noise is partially depended on PTA thresholds, but it may be more dependent on the abilities like frequency discrimination, frequency selectivity, and temporal resolution, compared to just the absolute thresholds.

**4.5 Correlation between SRTs and classification time thresholds.**

There was a significant positive correlation between SRTs and classification time thresholds for speech, noise, and music sounds. There was no correlation between SRT and classification time thresholds animal sounds. Therefore, the null hypothesis ($H_04$) that the classification time thresholds will not be correlated to SRT thresholds was only supported for animal sounds and not supported for speech, noise, and music sounds. This meant that the better a person was able to understand speech in noise, the shorter time it took for them to classify speech, noise, and music sounds.

As discussed in the Section 1.7, the glimpsing model of understanding speech in noise points out that the listeners extract information from short periods of temporal window where there is better signal to noise ratio (Kidd & Humes, 2012). Possessing a better ability to quickly recognise sounds of very brief duration may enable individuals to extract information even from shorter periods of temporal windows, allowing them to have more “glimpses”, compared to those with poorer temporal resolution. Therefore, they might end up with more information and better ability to recognise speech in noise compared to those with poorer temporal resolution of short sounds. Therefore, it could be assumed that the ability to quickly recognise different sound types of very short durations may in part help with the ability to understand speech in noisy environment.

A lot of factors are involved in both tests, it is difficult to conclude what factor are exactly responsible for the correlation between SRT and classification time thresholds. There are many processing abilities that are common in both tests. The short sounds of different classes used for testing the classification time thresholds contain complex information and requires complex processing which might involve auditory processing skills like temporal processing, perception of TFS, frequency discrimination and higher processing abilities like sound identification and classification. Good understanding of speech in noise requires excellent abilities for temporal processing, and frequency discrimination. Strelcyk and Dau, (2009) have also shown that perception of TFS is correlated to speech perception in noise.

**4.6 Relationship between age and hearing loss, and the correlation between age and classification time thresholds.**

The null hypothesis ($H_05$) that there will be no correlation between age and PTA thresholds was not supported. This meant that the higher the age of a participant the higher the PTA thresholds (worse the hearing). This is consistent with the previous finding of previous studies (Schuknecht, 1974; Schlauch & Nelson, 2015). The hearing loss as a result of aging is called presbycusis and it is one type of SNHL (Gates & Mills, 2005). This may be due to different sources of damage such as accumulated noise trauma and age-related damage of outer hair cells in the cochlea as well as structural and physiological changes in the peripheral and central structures (Tremblay & Ross, 2007). These changes and damages ultimately affect the signal transduction resulting in increased audiometric thresholds. Although, age correlated with PTA thresholds, it is very likely that the cause of SNHL was not limited to presbycusis in this study. Aging can cause a wide range of deficits in auditory system, resulting in numerous of deficits in processing abilities. However, the root cause of the hearing loss was not explored beyond the fact that they had SNHL.

The null hypothesis ($H_06$) that the classification time thresholds will not be correlated to age was not supported. There were significant correlations between the age and classification time thresholds for speech, noise, music, and animal sounds. This meant, that the older the participants the longer it took them to classify speech, noise, music, and animal sounds. Both age and hearing ability (PTA) were positively correlated with the classification time thresholds (ability to classify short sounds). The degree to which hearing loss and age independently contributed to performance of classifying short sounds were uncertain because aging can have same effect on perceptual as SNHL. Some deficits caused by aging which might have contributed to impaired performance in classifying sounds are also caused by hearing loss.

Some of the factors that are impacted by both aging and hearing loss which may be important in classifying short sounds are temporal processing, and sensitivity in perception of temporal fine structure. Studies have shown that temporal resolution measured by performance at gap detection tests be independently affected by both age and SNHL (Lister et al., 2011; Snell, 1997). Another factor affected by both age and SNHL may be perception of temporal fine structure. Hopkins and Moore, (2011) showed that the perception of temporal fine structure was poorer in older participants compared to younger participants even when both groups had similar hearing ability. Similarly, Strelcyk and Dau, (2009) showed that the perception of TFS is impaired in participants with SNHL.

There are also factors that are important in classification of short sounds that are not affected by aging but affected by hearing loss. Frequency selectivity which is important in the ability to distinguish vowels and musical instruments, is a likely factor involved in classification of short sounds. However, it is affected by hearing loss but not affected by aging. Hopkins and Moore, (2011) showed that the performance in the test of frequency selectivity was not significantly different for young and old participants with normal hearing. There are also factors that are affected by aging but not by hearing loss which are important in classification time thresholds such as attention and memory. See Section 1.8 for more details on the role of memory and attention in the auditory perception. Studies have shown that older participants performed poorly in tests that requires attention compared to younger participatns (Perbal, Droit-Volet, Isingrini, & Pouthas, 2002; Persad, Abeles, Zacks, & Denburg, 2002). The effect of aging on implicit memory is thought to be small but significant (Fleischman, Wilson, Gabrieli, Bienias, & Bennett, 2004). However, there are many other that have argued that age does not impact implicit memory. See Ward, Berry, and Shanks, (2013) for a review. Majority of studies have shown that explicit memory - the ability to consciously recall and recognise - is significantly affected by aging (Light, 1991; Davis, Trussell, & Klebe, 2001).

As mentioned above, there are many perceptual and cognitive factors involved in classification of short sounds such as perception of temporal fine structure, temporal processing, frequency selectivity, as well as attention and memory. Some of these factors are only affected by age or hearing loss and some are affected by both. Age and PTA thresholds were both correlated to classification time thresholds. Consequently, it could not be concluded to what degree the deficits related to age and hearing loss might have independently contributed to impaired performance in classifying short sounds.

**4.7 Practise effect for classification time thresholds with feedback.**

The null hypothesis ($H_0 7$) that there will be no significant difference between Run A and Run B of classification time thresholds was supported for speech, noise, and animal sounds but not for music sounds. This meant that participants did not do better in the second run compared to the first run for speech, noise, and animal sounds. However, classification time thresholds for music sounds with feedback required significantly shorter duration of sounds to classify music sounds in the second run (Run B) compared to the first run (Run A). In contrast, in the runs without feedback there was no significant difference between Run A and Run B for music sounds.

It has been shown that there is perceptual adaptation where participants learn to recognise certain acoustic cues like envelope or spectral cues properties to be associated with certain types of sounds (Shafiro, 2008; Gygi et al., 2004). This is particularly true when they are trained with feedback and cues. The current study gave participants feedbacks, which might have allowed them to associate the acoustic patterns of music to be associated to music sounds. It is not sure why the practice effect only occurred for music sounds and not in the other three sounds. However, as all the procedures for music was identical to that for all the other three sound classes, it could be assumed that the inherent properties of music sounds might have allowed the participants to learn a pattern. This might have provided an advantage in the subsequent run.

Music sounds have distinct organised acoustic patterns from which the listener derives the perception of melody, harmony, pitch, and timbre (Suied, Agus, Thorpe, Mesgarani, & Pressnitzer, 2014). The study by Gjerdingen and Perrott (2008), suggests that when classifying short music sounds to their genre, the participants most likely used perception of timbre of music rather than melody, harmony, or rhythm. However, there might have also been a limited amount of information on melody, harmony, and rhythm which aided in the classification. Timbre is the feature of sounds which allows discrimination between two sounds which have same duration pitch, and loudness (Krumhansl, 2010). Timbre consists of lots of acoustic cues and the spectrum of sound is one of the features of timbre. The distinct acoustic pattern of music and the timbre might have allowed participants to learn the patterns of music more easily compared to other sounds class when feedback was provided. The practice effect for music was not observed in separate runs without feedback. The practice effect across runs with feedback and runs without feedback was not analysed as the procedure. Instead, the effect of feedback was analysed across the two conditions.

**4.8 Effect of feedback on classification time thresholds.**

It was hypothesised ($H_0 8$) that there would be no significant difference between classification time thresholds with feedback and without feedback. The hypothesis was supported by the statistical analysis. Even though the addition of feedback had no impact on the classification time thresholds, it did impact the exclusion rate which was based on the absence of reversals and unrealistically high number of right responses as described in the Section 2.3.4. When the feedback was added in Part C, the number of data points excluded for speech, noise, and animal runs due to absence of reversals decreased by a significant amount. For noise runs, it decreased from 48.6% without feedback (Part B) to 27.9% with feedback (Part C). Similarly, for animal runs, from 20% to 10.3%, and for speech runs from 14.3% to 7.4%.

The feedback was initially removed after the pilot study, where it seemed that it was perpetuating the problem of participants classifying sounds based on durational differences between sounds classes, by acting as reinforcer. However, this change resulted in high number of exclusion rate in the Part B, therefore, the feedback was added again for Part C. The rationale for adding feedback back on was that the negative feedback might discourage people from constantly choosing one class by default when they were unsure. The rationale for the changes is described more thoroughly in the Method section.

Classification time thresholds measured with feedback were used instead of those measured without feedback for answering question in posed in Section 1.10. This is because the exclusion rate was lower in runs feedback and the feedback did not significantly impact the classification time thresholds. In addition, studies utilising alternative force choice, like the current study, are recommended to use feedback (Blackwell, 1952). Consequently, most studies with 4-AFC method do use feedbacks (Jäkel & Wichmann, 2006; Vancleef et al., 2018). Feedback seemed to provide the participants confidence that they were doing what they were supposed to do and prevented them from excessively choosing one option as a default response when they are unsure. Anecdotally, participants reported that they were more interested and motivated during feedback runs compared to the runs without feedback.

### 4.9 Application and clinical implications

This study also serves as an initial step towards reliably measuring classification time thresholds using logarithmic up-and-down procedure for participants with hearing loss and normal hearing. This study has detailed the problems it faced and ways it attempted to fix them, and how the changes affected the results. This would allow future studies to learn the problems that they may encounter during the testing of classification time thresholds. The areas where the future studies might need to improve for better measurement of classification time thresholds is outlined in Section 4.11.

Even though the studies on sound classification in humans are very hard to find, there are numerous studies that have developed software/algorithms for sound classifications that are incredibly fast (Lavner & Ruinskiy, 2009; Krishnamoorthy & Kumar, 2011; Yook et al., 2015). The classification time thresholds could also be used as a target for the algorithms to reach and exceed human capabilities. Although the current hearing aids use sound classifier to detect sound environments and reduce noise, their ability is limited and incomparable to human abilities. The use of an effective sound classifier in the hearing aids that can match the processing speed and abilities of human would be remarkable. It would not only be able to recognise different acoustic environments but also allow the hearing aids to selectively and effectively apply amplification to certain stream of sounds in the acoustic environment while ignoring others.

In addition, if we could predict the SRT using classification time thresholds, this test could serve as a simple and quick method for screening speech perception in noise via internet or smartphone, as calibration may not so critical. However, this would have to be checked using different presentation levels which can done by future studies.

**4.10 The exception: classification time thresholds for animal sounds**

Classification time thresholds for animal sound has been an exception in most of the analyses. There was no significant difference between classification time thresholds for animal sounds for participants with hearing impairment and with normal hearing but there was a significant difference for other sounds. There was significant positive correlation between pure tone thresholds and time classification thresholds for speech, noise, music but not for animal sounds. In addition, there was a significant positive correlation between SRT and all others except animal sounds.

It was not known why the response to animal sounds was completely different compared to others. The procedure, and testing equipment used were constant across all the sound types, so it was assumed that it must have either been the difference in stimuli, or in the perception and processing of animal sounds. However, the former reason seems more likely to be true than latter. Firstly, for all sound, the study aimed to diversify the source of sounds but keep them familiar to the listener, and one of the inclusion criteria for the sounds was that in should not have significant temporal gap. Finding 50 animal sounds that were familiar to NZ listener, and diverse with no significant temporal gap was challenging. As shown in the appendix B, the list of animal sound includes birds, mammals, insect, and even amphibians. This list needed to be balanced to include similar numbers for all sound groups. In this process, we might have included significant number of animal sounds which were unfamiliar to the participants. This might have caused the participants with hearing loss and normal hearing to equally perform in classifying animal sounds.

**4.11 Limitation and future studies.**

The study had few limitations. This was one of the first study that has used four types of sound for measuring classification time thresholds using logarithmic up-and-down procedure in participants with normal hearing and hearing loss. Therefore, there are many areas in the research that could be advanced and improved by future studies. First, we were unable to fix the problem causing the participants to score unrealistically high rate of correct responses, which resulted in high number of exclusion of data points. A possible reason was that the separation of one staircase from other three was providing durational cues. In the current procedure with four staircases simultaneously running, a natural separation of staircases in the working phase was observed. This is because one sound class may require shorter time to classify compared others. When the durational difference between the staircase is of perceptually different, there could have been durational cue for distinguishing the sound classes. With the feedback present would not have taken long for the participants to take advantage of these cues. In future study, increasing the proportion of the decoy trials of different lengths may alleviate the problem by reducing the occurrences of durational cues. However, a more effective way to prevent this might be to change the adaptive procedure, as the one used in the current study might be more prone to formation of durational cues with the natural separation of staircases. An adaptive procedure that produces staircase which fluctuates in terms of presentation of sound durations may be more averse to formation durational cues. Once such adaptive procedure is the A1 procedure developed by Brand and Kollmeier, (2002). Using A1 procedure requires less trials which makes it more efficient in comparison to the current procedure. The current procedure needs 20 trials before it gets close to the true thresholds and requires another 20 trials to fluctuate near the thresholds to build up reversals. Moreover, it is easier to monitor the accuracy using the A1 procedure as it calculates the thresholds by fitting psychometric curve to the data.

Second, the risk of including ambiguous sound stimuli could have been further reduced. Ambiguous sounds are sound of one class that could be perceived as another even at its full duration by participants with normal hearing. To reduce the risk of including ambiguous sounds in the stimuli set, each of the three researchers listened to it individually and removed the sounds that might sound ambiguous. However, each of the researcher could read the name or the source of the sound file before listening to it. Therefore, the researcher could be unintentionally biased towards it.

Thirdly, the study did not measure the familiarity of the sounds, as it assumed that the chosen sounds would be equally familiar to all the participants. Familiarity has been shown to be a strong influencer in perception of short sounds (Ballas, 1993; Shafiro, 2008).

In the future studies, one way to increase familiarity and reduce ambiguity would be to carry out a separate run, preceding the test of classification time thresholds. In these runs, all the untrimmed stimuli (around 500ms) would be played individually and participants would be asked to classify them as well as individually rate their familiarity (Murray et al., 2000). In the current study, this procedure was considered but not included due to time constraints and possible effect of fatigue on participants. Inclusion of this procedure would allow for the measurement of familiarity to the sound clips. Additionally, any untrimmed sound clips that were constantly and consistent misclassified as a specific sound class by multiple participants would be considered ambiguous and removed from the stimulus set. Reducing ambiguity is important as ambiguous sounds and can produce inaccurate results. For example, if the sound played is noise and most of the listeners hear it as music, then their "right" response would be record as being wrong. This would shift their true classification time threshold. One of the ways to spot presence of the ambiguous sounds, in the stimulus set, would be to look for consistent mistakes early on the trials. In beginning of each staircase, the sound clips are longer and mistakes in classification is not expected. The raw data was manually scanned to look for such mistakes at the beginning of each staircase. No significant and consistent mistakes were found indicating the lack of ambiguous sounds. However, the ambiguous sounds occurring later in the staircase, during shorter duration of presentation would be harder to spot. As the mistakes made in classification would be considered genuine.

Fourth, it was not sure how reliably the classification time thresholds measured the temporal resolution of the participants. One the reason for this is that the ability to classify may also require processing abilities like frequency discrimination, and frequency selectivity. Finding out to what degree each of these processing abilities affected the classification time thresholds was not within the scope of this study and may require significant changes to the procedure and the stimuli. The stimuli used in this study were not processed. It was the intention of the study to investigate the time required for classification of natural sounding sounds. The stimuli used in this study were from diverse sources, unprocessed, and contained complex acoustic features. Therefore, it was not known what exact acoustics features the participants used to classify the sounds. The relation between classification time thresholds and temporal resolution could be made more clearer, if a future study investigates the relationship between time classification thresholds and a well-recognised measure of temporal resolution like detection of gaps in noise (Moore, 2003). Additionally, the sound stimuli were manually analysed to look for any acoustic cues exclusive to one of sound types. For example, if noise sounds were always dominated by low frequency sounds, these cues were minimised by including noise sounds with high frequency noise. However, the study was unable to conclude the absolute absence of any distinct acoustic markers for a sound type that made it easier for them to recognise it in contrast to others.

In addition, the source of recruitment for participants could be diversified. Most of the participants with hearing loss were recruited from the university clinic and had significant hearing loss to requiring them to wear hearing aids. There were not as many people with mild and severe/profound hearing loss in the clinic records. They have been recruited from some other source. If the data for hearing loss was normally distributed, it would be expected that many other measures would also have normally distributed data. This is because the hearing loss correlates with other measures like age, SRT and classification time thresholds. Having a normal data would allow for more complex statistical analysis to be performed. It would also allow for a more precise, and stronger inference about the general population to be made (Tiemann, 2010).

**4.12 Summary of findings**

The classification time thresholds for speech, noise and music sounds was longer for participants with SNHL compared to those with normal hearing. There were significant correlations between classification time thresholds and pure tone average thresholds. The worse the hearing, the longer it took for the participants to classify speech, noise, and music sounds. Processing abilities like temporal resolution, perception of temporal fine structure, frequency selectivity, and frequency discriminations are all reduced in people with hearing loss (Glasberg et al., 1987; Zwicker & Schorn, 1978; Simon & Yund, 1993). These processing abilities are highly likely to be important for perception of short sounds. Therefore, impairment in these abilities may be one of the reasons that contributes to poorer performance in classifying of short sounds.

Speech sounds were classified faster in runs with feedback and without feedback by both hearing impaired and normal hearing participants. This finding is consistent with many others (Gray, 1942; Moradi et al., 2013; Isnard et al., 2016). The reason for faster processing of speech sounds may be that the human auditory system may be more efficient in processing speech sounds compared to all other sounds (Liberman et al., 1967; Moore, 2003).Some perceptual phenomena that are mostly present in speech perception have led some researchers to argue that for existence of a processing mode called "speech mode" and a special decoder (Liberman et al., 1967). However, some have doubted that they exist (Mehler & Hayes, 1981; Plomp, 2002).

There was significant correlation between PTA thresholds and the SRT measured by the UC Auditory-visual Matrix Sentence Test (UCAMST). The higher the degree of hearing loss, the harder it was for participants to understand speech in noise. This was consistent with other findings and the ability to understand speech in noise may be partial explained by the audiometric thresholds (Ching et al., 1998; Hopkins & Moore, 2011). However, multiple studies have suggested that the ability to understand speech in noise might be more dependent on the abilities affected by hearing loss like frequency discrimination, frequency selectivity, and temporal resolution compare to just the just the thresholds (Horst et al., 1990; Tyler et al., 1982; Glasberg & Moore, 1986)

The classification time thresholds for speech, noise and music sounds was positively correlated to SRTs. This might mean that the better ability in classifying sounds may aid in understanding speech in noise. The finding is consistent with glimpsing model of speech perception in noise (Kidd & Humes, 2012). Given this finding, the study suggested that better temporal resolution of short sounds may have enabled individuals with normal hearing to extract information even from shorter periods of temporal windows with better SNR. This might have allowed them to have more “glimpses”, compared to those with poorer temporal resolution.

There was a significant correlation between age and hearing loss which may have been partly due to presbycusis. There was also significant correlation between age and classification time thresholds for speech, noise, music, and animal sounds. The degree to which hearing loss and age independently contributed towards performance of classifying short sounds were uncertain because aging can have same effect on perceptual as SNHL. Some of the factors that are affected by both SNHL and aging are temporal resolution, and sensitivity in perception temporal fine structure (Snell, 1997; Lister et al., 2011; Hopkins & Moore, 2011).

There was no effect of feedback on classification time thresholds, but having feedback reduced the number of unrealistically high rates of correct responses and the lack of reversal in the working phase. There was no practise effect across Run A and Run B of classification time thresholds with feedback for speech, noise, and animal sounds but such effect was present for music sounds. The practise effect seen may be due to the inherent nature of music. This might have made it easier for the participants to recognise the music sounds more easily during the subsequent runs after having an initial exposure in the first run.

Classification time thresholds for animal sounds was an exception in most the analyses. It was not known why the responses to animal sounds were completely different. However, it was highly likely that animal sounds might have been unfamiliar to participants compared other sounds as it was very diverse. Familiarity has been known to affect classification of sounds (Murray et al., 2000). To solve this problem future studies might require the measurement of familiarity before testing sound classification.

This study has detailed the problems it faced, ways it attempted to fix them, and how the changes affected the results. This was one of the first study that used four types of sound for measurement classification time thresholds using logarithmic adaptive procedure of participants with normal hearing and hearing loss. Therefore, there were many areas in the research that could be advanced and improved by future studies. There were few limitations of the study. First, the inability to solve the problem of unrealistically high number of correct responses in the last 20 trials, particularly for noise sounds. Second, the study could have further reduced ambiguity. Third, the study could not measure familiarity. Fourth, the study was unable to confirm whether the classification of time thresholds measured the temporal resolution of the participants. In addition, the source of recruitment of participants could have been diversified and normal distributed data could have allowed for stronger ability to make inferences to the real world.

**4.13 Conclusions**

In conclusion, the study showed that the participants with hearing loss took longer to classify speech, noise, and music sounds. This may be due to impairment in processing abilities like temporal resolution, perception of temporal fine structure, frequency selectivity, and frequency discriminations. The study also showed that the better the ability in classifying short speech, noise, and music sounds, the better the understanding of speech in noise. This finding was consistent with glimpsing model of speech understanding in noise. Hearing ability was also correlated to ageing and was correlated to classification time thresholds. Both the effect of aging and hearing loss may cause deficits in abilities required for classifying short sounds. However, this study was separate the independent effect of age and SNHL on classification time thresholds. This study may serve as an initial step towards reliably measuring classification time thresholds for participants with hearing loss and normal hearing. This study has documented the procedures that are effective and those that are problematic. Some of the areas that could be improved are highlighted by this study. Many improvements could be made in the procedures and stimuli by future studies for more effective measurement of classification of short sounds.

## References


Agus, T. R., Suied, C., Thorpe, S. J., & Pressnitzer, D. (2012). Fast recognition of musical sounds based on timbre. *The Journal of the Acoustical Society of America, 131*(5), 4124-4133. doi:10.1121/1.3701865

Aibara, R., Welsh, J.T., Puria, S., Goode, R. L. (2001). Human middle-ear sound transfer function and cochlear input impedance. *Hearing Research 152*, 100–109.

Ashmore, J. F., & Kolston, P. J. (1994). Hair cell based amplification in the cochlea. *Current Opinion in Neurobiology, 4*(4), 503-508. doi:10.1016/0959-4388(94)90050-7

Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. Journal of Experimental Psychology: *Human Perception and Performance, 19(*2), 250-267. doi:10.1037/0096-1523.19.2.250

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature, 403*(6767), 309-312. doi:10.1038/3500207

Bethmann, A., & Brechmann, A. (2014). On the definition and interpretation of voice selective activation in the temporal cortex. *Frontiers in Human Neuroscience, 8(JULY)*, 499. doi:10.3389/fnhum.2014.00499

Blackwell, H. R. (1952). "Studies of psychophysical methods for measuring visual thresholds." *Journal of the Optical Society of America 42*(9): 606-616.

Blandy, S., Lutman, M. (2005) Hearing threshold levels and speech recognition in noise in 7-year-olds. *International Journal of Audiology 44*:435–443

Brand, T., & Kollmeier, B. (2002). Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests. *The Journal of the Acoustical Society of America, 111*(6), 2801-2810.

British Broadcasting Corporation (2018). *BBC Sound Effects BETA*. Retrieved from http://bbcsfx.acropolis.org.uk

Brokx, J., & Nooteboom, S. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics,10*(23).

Carhart, R., & Jerger, J. F. (1959). Preferred method for clinical determination of pure-tone thresholds. *Journal of Speech and Hearing Disorders, 24*(4), 330. doi:10.1044/jshd.2404.330

Ching, T., Dillon, H., & Byrne, D. (1998). Speech recognition of hearing-impaired listeners: Predictions from audibility and the limited role of high-frequency amplification. *Journal of Acoustical Society of America. 103*, 1128–1140.

Chiu, C. P., & Schacter, D. L. (1995). Auditory priming for nonverbal information: Implicit and explicit memory for environmental sounds. *Consciousness and Cognition, 4*(4), 440-458. doi:10.1006/ccog.1995.1050

Cooke, M. (2006). A glimpsing model of speech perception in noise. *Journal of the Acoustical Society of America*, 119(3), 1562. 10.1121/1.2166600

Cooper, N. P., & Rhode, W. S. (1995). Nonlinear mechanics at the apex of the guinea-pig cochlea. *Hearing Research, 82*(2), 225-243. doi:10.1016/0378-5955(94)00180-X

Cycowicz, Y. M., & Friedman, D. (1998). Effect of sound familiarity on the event-related potentials elicited by novel environmental sounds. *Brain and Cognition, 36*(1), 30-51. doi:10.1006/brcg.1997.0955

Dalton, D. S., Cruickshanks, K. J., Barbara, E. K., & Klein, R. (2003). The impact of hearing loss on quality of life in older adults. *The Gerontologist, 43*(5), 661-668. doi:10.1093/geront/43.5.661

Davis, H. P., Trussell, L. H., & Klebe, K. J. (2001). A ten-year longitudinal examination of repetition priming, incidental recall, free recall, and recognition in young and elderly. *Brain and Cognition. 46*, 99–104. doi: 10.1016/S0278-2626(01)80043-9

Digby, J.E., Purdy, S.C., Kelly, A.S. (2016). *Deafness Notification Report (2015) Hearing loss (not remediable by grommets) in New Zealanders under the age of 19.* Accessable: Auckland, New Zealand.

Duquesnoy, A.J. (1983). Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons. *Journal of the Acoustical Society of America 74*, 739–743.

Duthey, B. (2013). *Background Paper 6.21 Hearing Loss*. Geneva: WHO Int.

Ekdale, E. G. (2016). Form and function of the mammalian inner ear. *Journal of Anatomy, 228*(2), 324-337. doi:10.1111/joa.12308

Fleischman, D. A., Wilson, R. S., Gabrieli, J. D. E., Bienias, J. L., & Bennett, D. A. (2004). A longitudinal study of implicit and explicit memory in old persons. *Psychology and Ageing 19,* 617–625. doi: 10.1037/0882-7974.19.4.617

Fletcher H. (1940). Auditory patterns. Reviews of Modern Physics, 12, 47-65.

Fletcher, H., & Munson, W. A. (1933). Loudness, its definition, measurement and calculation. *Journal of the Acoustical Society of America 5,* 82–108.

Florentine, M., & Buus, S. (1984). Temporal gap detection in sensorineural and simulated hearing impairments. *Journal of Speech and Hearing Research, 27*(3), 449-455.

Florentine, M., Buus, S., Scharf, B., & Zwicker, E. (1980). Frequency selectivity in normally-hearing and hearing-impaired observers. *Journal of Speech and Hearing Research, 23*(3), 646.

Gates, G. A., & Mills, J. H. (2005). Presbycusis. *The Lancet, 366*(9491), 1111-1120. doi:10.1016/S0140-6736(05)67423-5

Giordano, B. L., McDonnell, J., & McAdams, S. (2010). Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds. *Brain and Cognition, 73*(1), 7-19. doi:10.1016/j.bandc.2010.01.005

Gjerdingen, R. O., & Perrott, D. (2008). Scanning the dial: The Rapid Recognition of Music Genres. *Journal of New Music Research, 37*(2), 93-100. doi:10.1080/09298210802479268

Glasberg, B. R., & Moore, B. C. J. (1992). Effects of envelope fluctuations on gap detection. *Hearing Research, 64*(1), 81-92. doi:10.1016/0378-5955(92)90170-R

Glasberg, B. R., Moore, B. C. J., & Bacon, S. P. (1987). Gap detection and masking in hearing-impaired and normal-hearing subjects. *The Journal of the Acoustical Society of America, 81*, 1546-1556.

Gomez, P., Ratcliff, R., & Perea, M. (2007). A model of the Go/No-go task. *Journal of Experimental Psychology: General, 136*(3), 389-413. doi:10.1037/0096-3445.136.3.389

Gordon-Salant, S., & Fitzgibbons, P. J. (1993). Temporal factors and speech recognition performance in young and elderly listeners. *Journal of Speech and Hearing Research, 36*, 1276-1285.

Gordon-Salant, S., & Fitzgibbons, P. J. (1999). Profile of auditory temporal processing in older listeners. *Journal of Speech, Language, and Hearing Research, 42,* 300-311.

Grant, K. W. (1987). Frequency modulation detection by normally hearing and profoundly hearing-impaired listeners. *Journal of Speech and Hearing Research, 30*(4), 558-563. 10.1044/jshr.3004.558

Gray, G. W. (1942). Phonemic Microtomy: *The Minimal Duration of Perceptible Speech Sounds. Communication Monograph 9*(1), 75-90.

Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *The Journal of the Acoustical Society of America, 115*(3), 1252-1265. doi:10.1121/1.1635840

Hagerman, B. (1982). Sentences for testing speech intelligibility in noise. *Scandinavian audiology, 11*(2), 79-87.

Hagerman, B., & Kinnefors, C. (1995). Efficient adaptive methods for measuring speech reception threshold in quiet and in noise. *Scandinavian Audiology, 24*(1), 71-77. doi:10.3109/01050399509042213

Hall, J. W., & Wood, E. J. (1984). Stimulus duration and frequency discrimination for normal-hearing and hearing-impaired subjects. *Journal of Speech and Hearing Research*, *27*(2), 252-256.

Haubert, N., & Pichora-Fuller, M. K. (1999). The perception of spoken language by elderly listeners: Contributions of auditory temporal processes. *Canadian Acoustics, 27*, 96-97

Healy, E. W., & Yoho, S. E. (2016). *Difficulty understanding speech in noise by the hearing impaired: Underlying causes and technological solutions.* Paper presented at the 89-92. 10.1109/EMBC.2016.7590647

Hochmuth, S., Brand, T., Zokoll, M. A., Castro, F. Z., Wardenga, N., & Kollmeier, B. (2012). A Spanish matrix sentence test for assessing speech reception thresholds in noise. *International journal of audiology, 51*(7), 536-544.

Hochmuth, S., Kollmeier, B., Brand, T., & Jürgens, T. (2015). Influence of noise type on speech reception thresholds across four languages measured with matrix sentence tests. *International Journal of Audiology, 54*(2), 62-70. doi:10.3109/14992027.2015.1046502

Hopkins, K., & Moore, B. C. J. (2011). The effects of age and cochlear hearing loss on temporal fine structure sensitivity, frequency selectivity, and speech reception in noise. *Journal of the Acoustical Society of America, 130*(1), 334-349. doi:10.1121/1.3585848

Horst, J. W., Javel, E., & Farley, G. R. (1990). Coding of spectral fine structure in the auditory nerve. II: Level-dependent nonlinear responses. *The Journal of the Acoustical Society of America, 88*(6), 2656-2681. doi:10.1121/1.399986

Isnard, V., Taffou, M., Viaud-Delmon, I., and Suied, C., (2016). Auditory Sketches: Very Sparse Representations of Sounds Are Still Recognizable. *Plos One.11*(3).

Jäkel, F., & Wichmann, F. A. (2006). Spatial four-alternative forced-choice method is the preferred psychophysical method for naïve observers. *Journal of Vision, 6*(11), 1307-13. doi:10.1167/6.11.13 11.

John, A. B., Hall, J. W., & Kreisman, B. M. (2012). Effects of advancing age and hearing loss on gaps-in-noise test performance. *American Journal of Audiology, 21*(2), 242-250. 10.1044/1059-0889(2012/11-0023)

Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Percept & Psychophysiology, 49*, 227-229.

Kalikow, D. N., Stevens, K. N., & Elliott, L. L. (1977). Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability. *The Journal of the Acoustical Society of America, 61*(5), 1337-1351. doi:10.1121/1.381

Kidd, G. R., & Humes, L. E. (2012). Effects of age and hearing loss on the recognition of interrupted words in isolation and in sentences. *The Journal of the Acoustical Society of America, 131*(2), 1434-1448. doi:10.1121/1.3675975

Killion, M. C., Niquette, P. A., Gudmundsen, G. I., Revit, L. J., & Banerjee, S. (2004). Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 116*(4), 2395-2405.

Klemmer, E.T. (1956). Time uncertainty in simple reaction time. *Journal of Experimental Psychology 51*,179–184

Kochkin, S. (2002). MarkeTrak VI: consumers rate improvements sought in hearing instruments. *Hearing Review 9,* 18–22

Krishnamoorthy, P., & Kumar, S. (2011). Hierarchical audio content classification system using an optimal feature selection algorithm. *Multimedia Tools and Applications,* 54(2), 415-444. doi:10.1007/s11042-010-0546-7

Krumhansl, C. L. (2010). Plink: "thin slices" of music. *Music Perception, 27*(5), 337-354. doi:10.1525/mp.2010.27.5.337

Lavner, Y., & Ruinskiy, D. (2009). A decision-tree-based algorithm for Speech/Music classification and segmentation. *Journal on Audio, Speech, and Music Processing*, *2009*(1), 1-14. doi:10.1155/2009/239892

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74*(6), 431-461. doi:10.1037/h0020279

Light, L. (1991). Memory and ageing: four hypotheses in search of data. *Annual Review of Psycholology. 42,* 333–376. doi: 10.1146/annurev.ps.42.020191.002001

Lister, J. J., Roberts, R. A., & Lister, F. L. (2011). An adaptive clinical test of temporal resolution: Age effects. *International Journal of Audiology, 50*, 367-374.

Lutman, M. E. (1991). Hearing disability in the elderly. *Acta Oto-Laryngologica. Supplementum 476,* 239–248.

McArdle, R., & Hnath-Chisolm, T. (2015). Speech Audiometry. In Katz, J. (Eds), *Handbook of Clinical Audiology. (p.p. 61-76).* Philadelphia, P.A: Wolters Kluwer Health

Mehler, J., & Hayes, R. W. (1981). The role of syllables in speech processing: Infant and adult data [and discussion]. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 295*(1077), 333-352.

Middelweerd, M. J., Festen, J. M., Plomp, R. (1990). Difficulties with speech intelligibility in noise in spite of a normal pure-tone audiogram. *Audiology 29*:1–7.

Moore B.C.J, & Glasberg, B. R. (1986). Comparisons of frequency selectivity in simultaneous and forward masking for subjects with unilateral cochlear impairments. *Journal of the Acoustical Society of America*, *80*, 93-107.

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, *9*(4), 399-406. doi:10.1007/s10162-008-0143-x

Moore, B. C. J., Peters, R. W., & Glasberg, B. R. (1996). Detection of decrements and increments in sinusoids at high overall levels. *Journal of the Acoustical Society of America, 99*(6), 3669-3677. 10.1121/1.414964

Moore, B. C. J., Wojtczak, M., & Vickers, D. A. (1996). Effect of loudness recruitment on the perception of amplitude modulation. *The Journal of the Acoustical Society of America, 100*(1), 481-489. doi:10.1121/1.415861

Moore, B. C., Glasberg, B. R., Plack, C. J., & Biswas, A. K. (1988). The shape of the ear's temporal window. *The Journal of the Acoustical Society of America, 83*(3), 1102.

Moore, B.C. J., (2013). *An Introduction to Psychology of Hearing* (6th ed). Bingley, United Kingdoms: Emerald Group Publishing Limited

Moradi, S., Lidestam, B., Rönnberg, J., (2013). Gated audiovisual speech identification in silence vs. noise: Effects on time and accuracy. *Frontiers in Psychology, 4*(359), 359. doi:10.3389/fpsyg.2013.00359

Mueller, H. (2003). Fitting test protocols are "more honoured in the breach than the observance." *Hearing Journal 56:*19–26.

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain identification of sounds of objects. *Journal of Neuroscience. 26,* 1293–1302.

Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2000). Rapid Brain Discrimination of Sounds of Objects. *Journal of Neuroscience. 26*(4), 1293-1302.

National Institute of Heath (2019). *NIDCD Fact Sheet: Age-Related Hearing loss.* Retrieved from https://www.nidcd.nih.gov/sites/default/files/Documents/health/hearing/AgeRelatedHearingLoss.pdf

New Zealand Audiological Society, (2007). *New Zealand Audiological Society Best Practice Guideline: adult pure-tone audiometry.* Retrived from https://www.audiology.org.nz/UserFiles/file/_private/adult%203.pdf

Nilsson, M., Soli, S. D., & Sullivan, J. A. (1994). Development of the hearing in noise test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America, 95*(2), 1085-1099. doi:10.1121/1.408469

Obert, S., Tchorz, J., (2018). *Geräuchklassifikation.* Unpublished manuscript, Lübeck University of Applied Science, Lübeck, Germany.

Oxenham A. J., & Moore, B. C.J. Modeling the additivity of nonsimultaneous masking. Hearing Research, 80,105-118.

Ogg, M., Slevc, L. R., & Idsardi, W. J. (2017). The time course of sound category identification: Insights from acoustic features. *The Journal of the Acoustical Society of America, 142*(6), 3459-3473. doi:10.1121/1.5014057

Ozimek, E., Warzybok, A., & Kutzner, D. (2010). Polish sentence matrix test for speech intelligibility measurement in noise. *International journal of audiology, 49*(6), 444-454.

Patterson, R.D., & Moore, B.C.J. (1986). Auditory filters and excitation pattern as representation of frequency resolution. In B.C.J. Moore (Ed.), *Frequency selectivity in Hearing*. London. Academic Press.

Patuzzi, R. (2009). Cochlear Mechanics *Encyclopedia of Neuroscience* (pp. 1041-1049). Oxford: Academic Press

Perbal, S., Droit-Volet, S., Isingrini, M., & Pouthas, V. (2002). Relationships between age-related changes in time estimation and age-related changes in processing speed, attention, and memory. *Aging, Neuropsychology, and Cognition, 9*(3), 201-216. doi:10.1076/anec.9.3.201.9609

Peretz, I., Gagnon, L., & Bouchard, B. (1998). Music and emotion: Perceptual determinants, immediacy, and isolation after brain damage. *Cognition*, *68,* 111–141.

Persad, C. C., Abeles, N., Zacks, R. T., & Denburg, N. L. (2002). Inhibitory changes after age 60 and their relationship to measures of attention and memory. *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences, 57*(3), 223-232.

Phillips, S. L., Gordon-Salant, S., Fitzgibbons, P. J., & Yeni-Komshian, G. (2000). Frequency and temporal resolution in elderly listeners with good and poor word recognition. *Journal of Speech, Language, and Hearing Research*, *43*, 217-228.

Pichora-Fuller, M. K., Schneider, B. A., MacDonald, E., Pass, H., & Brown, S. (2007). Temporal jitter disrupts speech intelligibility: *A simulation of auditory aging. Hearing Research*, 223, 114-121.

Pick, G., Evan, E. F., & Wilson, J. P. (1977). Frequency resolution in patients with hearing loss of cochlear origin. In E. F. Evans & J.P. Wilson (Eds.), *Psychophysics and Physiology of Hearing.* London: Academic Press.

Pickles, J. O. (2012). *An Introduction to the Physiology of Hearing* (4th ed.). United Kingdom: Emerald Group Publishing Limited.

Plomp, R. (2002). *The intelligent ear.* Erlbaum, Mahwawj, NJ

Plomp, R., & Mimpen, A. (1979). Improving the reliability of testing the speech reception threshold for sentences. *International Journal of Audiology, 18*(1), 43-52.

Puglisi, G. E., Warzybok, A., Hochmuth, S., Astolfi, A., Prodi, N., Visentin, C., & Kollmeier, B. (2014). *Construction and first evaluation of the Italian Matrix Sentence Test for the assessment of speech intelligibility in noise*. Paper presented at the PROCEEDINGS OF FORUM ACUSTICUM.

Rice, J.J., May, B.J., Spirou, G.A., & Young, E.D. (1992). Pinna-based spectral cues for sound localization in cat. *Hearing Research 58*, 132–152.

Rigoulot, S., Pell, M. D., & Armony, J. L. (2015). Time course of the influence of musical expertise on the processing of vocal and musical sounds. *Neuroscience, 290*, 175-184. doi:10.1016/j.neuroscience.2015.01.0033

Ruggero, M.A., & Temchin, A.N. (2002). The roles of the external, middle, and inner ears in determining the bandwidth of hearing. *Preceding of National Academy of Science USA 99*, 13206–13210.

Scarinci, N., Worrall, L., & Hickson, L. (2008). The effect of hearing impairment in older people on the spouse. *International Journal of audiology*, *47*(3), 141-151.

Schlauch, R. S., & Nelson, P. (2015). Puretone Evaluation. In Katz, J. (Eds), *Handbook of Clinical Audiology. (p.p. 28-48).* Philadelphia, P.A: Wolters Kluwer Health

Schneider, B. A., & Pichora-Fuller, M. K. (2001). Age-related changes in temporal processing: Implications for speech perception. *Seminars in Hearing, 22*, 227-240.

Schneider, B. A., Pichora-Fuller, M. K., Kowalchuk, D., & Lamb, M. (1994). Gap detection and the precedence effect in young and old adults. *The Journal of the Acoustical Society of America, 95*, 980-991.

Schuknecht, H. F., (1974). *Pathology of Ear*. Cambridge, MA, Harvard University Press.

Shafiro, V. (2008). Development of a large-item environmental sound test and the effects of short-term training with spectrally-degraded stimuli. *Ear and Hearing, 29*(5), 775-790. doi:10.1097/AUD.0b013e31817e08ea

Shen, D., Vuvan, D. T., & Alain, C. (2018). Cortical sources of the auditory attentional blink. *Journal of Neurophysiology, 120*(2), 812.

Simon, H. J., & Yund, E. W. (1993). Frequency discrimination in listeners with sensorineural hearing loss. *Ear and Hearing, 14*(3), 190-201. 10.1097/00003446-199306000-00006

Smoorenburg, G. (1992) Speech reception in quiet and in noisy conditions by individuals with noise-induced hearing loss in relation to their tone audiogram. *The Journal of the Acoustical Society of America,91*: 421–437.

Snell, K. B. (1997). Age-related changes in temporal gap detection. *The Journal of the Acoustical Society of America, 101*, 2214-2220.

Snell, K. B., & Frisina, D. R. (2000). Relationships among age-related differences in gap detection and word recognition. *The Journal of the Acoustical Society of America, 107,* 1615-1626.

Snell, K. B., Mapes, F. M., Hickman, E. D., & Frisina, D. R. (2002). Word recognition in competing babble and the effects of age, temporal processing, and absolute sensitivity. *The Journal of the Acoustical Society of America, 112*(2), 720-727. doi:10.1121/1.1487841

Stasiunas, A., Verikas, A., Kemesis, P., Bacauskiene, M., Miliauskas, R., Stasiuniene, N., & Malmqvist, K. (2003). A non-linear circuit for simulating OHC of the cochlea. *Medical Engineering and Physics, 25*(7), 591-601. doi:10.1016/S1350-4533(03)00071-7

Strelcyk, O., & Dau, T. (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *Journal of Acoustical Society of America*, *125,* 3328–3345.

Strouse, A., Ashmead, D. H., Ohde, R. N.,& Grantham, D.W. (1998). Temporal processing in the aging auditory system. *The Journal of the Acoustical Society of America, 104*, 2385-2399.

Suied, C., Agus, T. R., Thorpe, S. J., Mesgarani, N., & Pressnitzer, D. (2014). "Fast recognition of musical sounds based on timbre," *The Journal of the Acoustical Society of America, 135*, 1380–1391.

Tambs, K. (2004). Moderate effects of hearing loss on mental health and subjective well-being: Results from the nord-trøndelag hearing loss study. *Psychosomatic Medicine, 66*(5), 776

Thiesen, F, C., Kopiez, R., Reuter, C., zedik-Eysenberg, I., & Schlemmer, K., (2016). In *the Blink of an Ear: A Critical Review of Very Short Musical Elements (Plinks)*, presented at ICMPC14, July 5-9, 2016, San Francisco, ResearchGate.

Tiemann, T. K., (2010). *Introductory Business Statistics With Interactive Spreadsheets*. Retrieved from https://opentextbc.ca/introductorybusinessstatistics/

Tierstimmenarchiv, (2018). *Animal Sound Archive of Museum für Naturkunde, Berlin.* Retrieved from http://www.animalsoundarchive.org

Tremblay, K., & Ross, B. (2007). Effects of age and age-related hearing loss on the brain. *Journal of Communication Disorders, 40*(4), 305-312.

Trounson, R. H. (2012). *Development of the UC Auditoryvisual Matrix Sentence Test. MAud Thesis.* Department of Communication Disorders. The University of Canterbury.

Turner, C. W., Chi, S., & Flock, S. (1999). Limiting spectral resolution in speech for listeners with sensorineural hearing loss. *Journal of Speech, Language, and Hearing Research, 42(*4), 773-784. doi:10.1044/jslhr.4204.773

Turner, C. W., Fabry, D. A., Barrett, S., & Horwitz, A. R. (1992). Detection and recognition of stop consonants by normal hearing and hearing-impaired listeners. *Journal of Speech and Hearing Research, 35*(4), 942-949. 10.1044/jshr.3504.942

Tyler, R. S., Summerfield, Q.,Wood, E. J.,& Fernandes, M. A. (1982). Psychoacoustic and phonetic temporal processing in normal and hearing-impaired listeners. *The Journal of the Acoustical Society of America, 72*, 740-752.

Tyler, R.S. (1986). Frequency resolution in hearing impaired listeners. In B.C.J. Moore (Eds.), *Frequency Selectivity in Hearing,* London; Academic Press.

Vancleef, K., Read, J. C. A., Herbert, W., Goodship, N., Woodhouse, M., & Serrano-Pedraza, I. (2018). Two choices good, four choices better: For measuring stereoacuity in children, a four-alternative forced-choice paradigm is more efficient than two. *PloS One, 13*(7), e0201366. doi:10.1371/journal.pone.0201366

Wagener, K. C., Brand, T., & Kollmeier, B. (June, 2007). *International cross-validation of sentence intelligibility tests*. Paper presented at Congress of the German Society of Audiology. Heidelberg, Germany. Retrieved from https://www.uzh.ch/orl/dga2007/program/scientificprogram/Wagener__K.C._et_al.pdf

Wagener, K., & Brand, T. (2005). Sentence intelligibility in noise for listeners with normal hearing and hearing impairment: Influence of measurement procedure and masking parameters. *International Journal of Audiology*, *44*(3), 144-156. doi: 10.1080/14992020500057517

Ward, E. V., Berry, C. J., & Shanks, D. R. (2013). Age effects on explicit and implicit memory. Frontiers in Psychology, 4, 639. doi:10.3389/fpsyg.2013.00639

Wilson, R. H., McArdle, R. A., & Smith, S. L. (2007). An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *Journal of Speech, Language, and Hearing Research, 50*(4), 844-856.

Wolmetz, M., & Elhilali, M. (2016). Attentional and contextual priors in sound perception. *PloS One, 11*(2), e0149635. doi:10.1371/journal.pone.0149635

Wong, A. C. Y., & Ryan, A. F. (2015). Mechanisms of sensorineural cell damage, death and survival in the cochlea. *Frontiers in Aging Neuroscience, 7,* 58. 10.3389/fnagi.2015.00058

Yook, S., Nam, K. W., Kim, H., Hong, S. H., Jang, D. P., & Kim, I. Y. (2015). An environment-adaptive management algorithm for hearing-support devices incorporating listening situation and noise type classifiers: Management algorithm for hearing-support devices. *Artificial Organs, 39*(4), 361-368. doi:10.1111/aor.12391

Zurek, P. M., & Formby, C. (1981). Frequency-discrimination ability of hearing-impaired listeners. *Journal of Speech and Hearing Research, 24(*1), 108-112. 10.1044/jshr.2401.108

Zwicker, E., & Schorn, K. (1978). Psychoacoustical tuning curves in audiology. *Audiology: Official Organ of the International Society of Audiology, 17*(2), 120.

## Appendix A

- Human ethics Approval letter
- Invitation Letter for Part A and Part B
- Information Sheet for Part A
- Invitation Letter for Part C
- Information Sheet for Part B and C
- Consent form

- Human ethics Approval letter

UNIVERSITY OF CANTERBURY
*Te Whare Wānanga o Waitaha*
CHRISTCHURCH NEW ZEALAND

HUMAN ETHICS COMMITTEE

Secretary, Rebecca Robinson
Telephone: +64 03 369 4588, Extn 94588
Email: human-ethics@canterbury.ac.nz

Ref: HEC 2018/33/LR

19 June 2018

Durga Lal Budathoki
Communication Disorders
UNIVERSITY OF CANTERBURY

Dear Durga

Thank you for submitting your low risk application to the Human Ethics Committee for the research proposal titled "Duration Thresholds for Identifying Different Sound Types".

I am pleased to advise that this application has been reviewed and approved.

Please note that this approval is subject to the incorporation of the amendments you have provided in your email of 13th June 2018.

With best wishes for your project.

Yours sincerely

*pp.* R. Robinson

Professor Jane Maidment
***Chair, Human Ethics Committee***

- Invitation Letter for Part A and Part B

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

VOLUNTEERS NEEDED

Dear all,

We would like to invite you to participate in a research project at the University of Canterbury Speech and Hearing Clinic.

We are investigating a new way of predicting people's ability to understand speech in noise using a relatively short and simple test: measuring the shortest time required to identify sound sounds types. A summary of research procedures has been attached along with this letter along with the consent form.

The research will be held in the University of Canterbury speech and Hearing Clinic in late September and throughout October. The experiment will take around 1 hour and 15 Minutes. You will:

- Receive free hearing check
- Get involved in development of new and quick way of predicting speech understanding in noise
- Receive a $10 petrol voucher

If you are interested, or would like more information, please contact **Durga Lal Budathoki** at **durgalal.budathoki@pg.canterbury.ac.nz**, or supervisor **Greg O'Beirne** at **gregory.obeirne@canterbury.ac.nz**

Thank you,

Durga Lal Budathoki,
Masters of Audiology (2nd year student)

*This project has been reviewed and approved by the University of Canterbury Human Ethics Committee*

- Information Sheet for Part A

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

## Duration thresholds for identifying different sound types

**Information Sheet for Persons Participating in the Research studies**

**Primary Researchers**: Durga Lal Budathoki (MAud Student, 2nd Year)

**Research Supervisor:** Associate Professor Greg O'Beirne, University of Canterbury

**Research Supervisor:** Professor Jürgen Tchorz, Luebeck University of Applied Science

The aim of the study is to find a way to predict the speech understanding in noise using relatively short and simple tests that measures time required for an individual to classify the sounds types.

To be eligible to participate in the study you must:

- be 18 years old or older
- be a native New Zealand English speaker
- have no significant memory problems or dexterity issues
- have no current ear infections

If you choose to take part in this study, firstly, we will look in the ear. You will then have a hearing test (called "pure tone audiometry", where you will be given a button to push whenever you hear a tone), to measure the quietest level of sound you can detect. The sounds will be delivered to the ear by small foam earphones or headphones. This should take around 30 minutes to complete.

After this, if the hearing test thresholds are within our inclusion criteria, you will take part in a "speech in noise" test. This is a short test that will present speech sentences along with noise of different intensities within comfortable listening levels. This will measure your ability to understand speech in different noise levels. This will take around 10 minutes.

Finally, you will perform the tests to determine your classification time thresholds for speech sounds. You will be presented with 160 short sound clips of 4 different types: speech, noise, animal sounds, and music. Each time a clip is played, you will be instructed to choose one of the 4 options from the screen. The screen will provide feedback by revealing your answer as right or wrong. The duration of this clips will vary to find the minimum time duration required for each types of sound to be identified. This should take around 15 minutes.

Participation is voluntary and you have the right to withdraw at any stage without penalty or explanation. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, only your initials and identification number will

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

be recorded in the data. The data will be kept in a secure locked facility or password protected electronic form. A thesis is a public document and will be available through the UC Library.

Please indicate on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out as a requirement for the Master of Audiology degree of Durga Lal Budathoki under the supervision of Greg O'Beirne and Jürgen Tchorz. We will be pleased to discuss any concerns you may have about participation in the project.

**Durga Lal Budathoki (2nd year MAud Student)**
Masters of Audiology, Department of Communication Disorders
Email: durgalal.budathoki@pg.canterbury.ac.nz

**Professor Jürgen Tchorz,**
**Secondary research supervisor**
Luebeck University of Applied Science,
Lubeck, Germany
Email: tchorz@fh-luebeck.de
Telephone No : +49 451 300 5240

**Greg O'Beirne, PhD**
**Primary research supervisor & Associate Professor in Audiology**
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: gregory.obeirne@canterbury.ac.nz
Telephone No : +6433694313

If you have any complaints it should address be addressed to The Chair of Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz). Telephone No: +64 3 364 2987.

- Invitation Letter for Part C

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

VOLUNTEERS NEEDED

Dear all,

We would like to invite you to participate in a research project at the University of Canterbury Speech and Hearing Clinic.

We are investigating a new way of predicting people's ability to understand speech in noise using a relatively short and simple test: measuring the shortest time required to identify sound sounds types. A summary of research procedures has been attached along with this letter along with the consent form.

The research will be held in the University of Canterbury speech and Hearing Clinic in late September and throughout October. The experiment will take around 1 hour and 15 Minutes. You will:

- Receive free hearing check
- Get involved in development of new and quick way of predicting speech understanding in noise
- **Receive a $20 petrol voucher**

If you are interested, or would like more information, please contact **Durga Lal Budathoki** at **durgalal.budathoki@pg.canterbury.ac.nz**, or supervisor **Greg O'Beirne** at **gregory.obeirne@canterbury.ac.nz**

Thank you,

Durga Lal Budathoki,
Masters of Audiology (2nd year student)

*This project has been reviewed and approved by the University of Canterbury Human Ethics Committee*

- Information Sheet for Part B and C

UC UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

## Duration thresholds for identifying different sound types

**Information Sheet for Persons Participating in the Research studies**

**Primary Researchers**: Durga Lal Budathoki (MAud Student, 2nd Year)

**Research Supervisor:** Associate Professor Greg O'Beirne, University of Canterbury

**Research Supervisor:** Professor Jürgen Tchorz, Luebeck University of Applied Science

The aim of the study is to find a way to predict the speech understanding in noise using relatively short and simple tests that measures time required for an individual to classify the sounds types.

To be eligible to participate in the study you must:

- be 18 years old or older
- be a native New Zealand English speaker
- have no significant memory problems or dexterity issues
- have no current ear infections

If you choose to take part in this study, firstly, we will look in the ear. You will then have a hearing test (called "pure tone audiometry", where you will be given a button to push whenever you hear a tone), to measure the quietest level of sound you can detect. The sounds will be delivered to the ear by small foam earphones or headphones. This should take around 30 minutes to complete.

After this, if the hearing test thresholds are within our inclusion criteria, you will take part in a "speech in noise" test. This is a short test that will present speech sentences along with noise of different intensities within comfortable listening levels. This will measure your ability to understand speech in different noise levels. This will take around 10 minutes.

Finally, you will perform the tests to determine your classification time thresholds for speech sounds. You will be presented with 200 short sound clips of 4 different types: speech, noise, animal sounds, and music. Each time a clip is played, you will be instructed to choose one of the 4 options from the screen. The screen will provide feedback by revealing your answer as right or wrong. The duration of this clips will vary to find the minimum time duration required for each types of sound to be identified. This should take around 20 minutes.

Participation is voluntary and you have the right to withdraw at any stage without penalty or explanation. You may ask for your raw data to be returned to you or destroyed at any point. If you withdraw, I will remove information relating to you.

The results of the project may be published, but you may be assured of the complete confidentiality of data gathered in this investigation: your identity will not be made public without your prior consent. To ensure anonymity and confidentiality, only your initials and identification number will

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

be recorded in the data. The data will be kept in a secure locked facility or password protected electronic form. A thesis is a public document and will be available through the UC Library.

Please indicate on the consent form if you would like to receive a copy of the summary of results of the project.

The project is being carried out as a requirement for the Master of Audiology degree of Durga Lal Budathoki under the supervision of Greg O'Beirne and Jürgen Tchorz. We will be pleased to discuss any concerns you may have about participation in the project.

**Durga Lal Budathoki (2nd year MAud Student)**
Masters of Audiology, Department of Communication Disorders
Email: durgalal.budathoki@pg.canterbury.ac.nz

**Professor Jürgen Tchorz,**
**Secondary research supervisor**
Luebeck University of Applied Science,
Lubeck, Germany
Email: tchorz@fh-luebeck.de
Telephone No : +49 451 300 5240

**Greg O'Beirne, PhD**
**Primary research supervisor & Associate Professor in Audiology**
Department of Communication Disorders
University of Canterbury
Private Bag 4800, Christchurch 8140, New Zealand
Email: gregory.obeirne@canterbury.ac.nz
Telephone No : +6433694313

If you have any complaints it should address be addressed to The Chair of Human Ethics Committee, University of Canterbury, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz). Telephone No: +64 3 364 2987.

- Consent form for all

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

## Duration thresholds for identifying different sound types

**Consent Sheet for Persons Participating in the Research studies**

- ☐ I have been given a full explanation of this project and I have been given an information sheet. I have had the opportunity to ask questions
- ☐ I understand what is required of me if I agree to take part in the research.
- ☐ I understand that participation is voluntary and I may withdraw at any time without penalty. Withdrawal of participation will also include the withdrawal of any information I have provided should this remain practically achievable.
- ☐ I understand that any information or opinions I provide will be kept confidential to the researchers mentioned in the information sheet and that any published or reported results will not identify the participants. I understand that a thesis is a public document and will be available through the UC Library.
- ☐ I understand that all data collected for the study will be kept in locked and secure facilities and/or in password protected electronic form and will be destroyed after *five*/years.
- ☐ I understand the risks associated with taking part and how they will be managed.
- ☐ I understand that I can contact the researcher Durga lal Budathoki (durgalal.budathoki@pg.canterbury.ac.nz) or supervisor Greg O'Beirne (gregory.obeirne@canterbury.ac.nz). If I have any complaints, I can contact the Chair of the University of Canterbury Human Ethics Committee, Private Bag 4800, Christchurch (human-ethics@canterbury.ac.nz).
- ☐ I would like a summary of the results of the project.
- ☐ By signing below, I agree to participate in this research project.

Name:____________________________Signed:__________________________Date:________________

Email address *(for report of findings, if applicable)*

______________________________________________________________________________

*Please return the consent form to the researcher before participating in the study.*

## Appendix B

- List of names of sound files used for trials
- List of names of sound files used as decoys

| Sound types | Speech | Noise | Animal | Music |
|---|---|---|---|---|
| 1. | audiobook.wav | High Wind In Bushes, occasional bird song..wav | African buff.wav | instru ad.wav |
| 2. | Australian women3.wav | Airplane Take Off Sound Effect In High Quality.wav | Alligator.wav | instru beats.wav |
| 3. | Boy1.wav | AMSR.wav | Alopex.wav | instru bollywood.wav |
| 4. | Elderlyman.wav | Ball and chain demolition in progress with crane manoeuvring.wav | antelope.wav | instru classical.wav |
| 5. | Female 3.wav | beer garden people talking.wav | baboon2.wav | instru country.wav |
| 6. | Female 4.wav | boat on water.wav | bear cat.wav | instru disco.wav |
| 7. | Female 6.wav | bonfire.wav | bird2.wav | instru electro.wav |
| 8. | Female 7 good.wav | Car starting.wav | bird3.wav | Instru filmscore.wav |
| 9. | female good quality.wav | Christchurch tram.wav | bird4.wav | instru folk instru.wav |
| 10. | Female news.wav | clapss.wav | bird9.wav | instru funk.wav |
| 11. | Female1.wav | cloth AMSR.wav | bird10.wav | instru latin.wav |
| 12. | Female2.wav | drilling.wav | birds5.wav | instru mellow 2.wav |
| 13. | Female5.wav | elcetrical monotony.wav | Birds6.wav | instru new age ambigious.wav |
| 14. | Girl1.wav | emptying sink.wav | Birds7.wav | instru world.wav |
| 15. | guy.wav | engine 2 strokes .wav | Birds8.wav | instruJaaz.wav |
| 16 | male from news.wav | Football crowds and claps.wav | cat.wav | instrum classical.wav |
| 17. | Male1.wav | glass broken.wav | chimp.wav | instrumental kids.wav |
| 18. | male2.wav | helicopter.wav | Cow.wav | instrumental news music uplift good quality.wav |

| | | | | |
|---|---|---|---|---|
| 19. | Male3.wav | high frequency radio noise.wav | Cricket Sound.wav | instrumental up beat.wav |
| 20. | male4.wav | high pitch Morse code.wav | crocodile.wav | instrumental.wav |
| 21. | male5.wav | High pitched wind.wav | dog.wav | modern guitar and singing .wav |
| 22. | man good .wav | High-pressure air pipes singing..wav | donkey.wav | modern non-instru.wav |
| 23. | mannz1.wav | lifts.wav | elephant.wav | music instru good.wav |
| 24. | mannz2.wav | machine sounds.wav | fowl.wav | music instru.wav |
| 25. | mannz3.wav | noise and bikes rining.wav | Frog.wav | non gospel.wav |
| 26. | news man 2.wav | noisy apple eater.wav | goat.wav | non instru 1.wav |
| 27. | news man.wav | non instrumental: fire.wav | goose.wav | non instru 2.wav |
| 28. | news music instru, good.wav | outdoor haircut.wav | Guinee pig.wav | non instru children.wav |
| 29. | newswomen.wav | paper folding.wav | Horse1.wav | non instru country.wav |
| 30. | school boy .wav | Pedestrianized high street, with passing footsteps, chatter and children..wav | Hyla_arborea.wav | non instru pop.wav |
| 31. | school boy 3.wav | push lawnmower.wav | jackle.wav | non instru RnB.wav |
| 32. | school girl3.wav | River.wav | lemur.wav | non-instru .wav |
| 33. | school girl4.wav | rollercoster.wav | leopard.wav | non-instru musical.wav |
| 34. | school girl5.wav | small stream.wav | lion.wav | non-instrupop.wav |
| 35. | school girls 1.wav | theater claps.wav | red deer.wav | noninstru crimson rose.wav |
| 36. | school girls 2.wav | tree falling into river.wav | rhino.wav | noninstru world.wav |
| 37. | schoolboy 2.wav | typing.wav | Roosters.wav | rap2.wav |
| 38. | toddler1.wav | weaving.wav | seal.wav | REGGAE.wav |
| 39. | women good quality.wav | Windsurfing, board launched from gravel beach.wav | sheep1.wav | Rock.wav |
| 40. | women good.wav | whistle and crowds.wav | wolf.wav | soul noninstru.wav |

- List of names of sound files used as decoys

| Sound types | Speech | Noise | Animal | Music |
|---|---|---|---|---|
| 1. | american guy.wav | AMSR2.wav | baboon.wav | blues.wav |
| 2. | boyfrom news.wav | bells.wav | bird 4.wav | instru 2.wav |
| 3. | man .wav | car racing.wav | bird.wav | instru classic.wav |
| 4. | man2.wav | chains on buckets.wav | bird1.wav | instru dance.wav |
| 5. | News man.wav | explosion.wav | cat copy.wav | instru mellow.wav |
| 6. | news women.wav | falling on glass table.wav | cat2.wav | instru1.wav |
| 7. | nz accent man.wav | lifts .wav | Cicada - New Zealand.wav | rap.wav |
| 8. | nz guy.wav | projectors.wav | cow3.wav | vocal 2.wav |
| 9. | nz women2.wav | steam.wav | Kiwi(500ms).wav | vocal 90s.wav |
| 10. | nz women3.wav | traffic noise india.wav | snow cat.wav | wordy music nz low quality.wav |